

China's hopes and hypes

The scientific potential of China is great. Recent initiatives reflect the government's justified ambitions for research. They also highlight unjustified secrecy and misguided policy agendas.

Scientific research in China stands to benefit greatly from a recent push by the government. Since the beginning of this year, the government has been piecing together a major initiative in nanotechnology. Drawing on the momentum of the Human Genome Project, the possibility of sequencing various other genomes is being investigated, and post-genomic projects are becoming widespread. Universities, especially those of Peking and Tsinghua, home to a recent large-scale biochip initiative, are being developed as hubs of collaboration between industry and academia. Just two weeks ago, the Ministry of Science and Technology announced a 15 billion yuan (US\$1.8 billion) investment in civil programmes for research and development. This will continue the agenda of the "863 Program" — China's first high-tech R&D plan, announced in March 1986 — which targeted eight scientific disciplines, including the biosciences, information technology and aeronautics.

But there is a danger. Close government involvement, without an adequate external review system, could lead to fruitless projects and could even stall scientific development. The risk is that big projects will become an end in themselves, where both politicians and scientists demonstrate their power by the size of the budget they acquire for their project, regardless of its scientific merit. A desire for big projects is not uncommon. But in China many decisions about budgets and new projects are made in back rooms, so that people often do not know the justification for the choice of one programme over another. Researchers with experience of both countries compare China unfavourably to the United States in transparency: China's decision-making processes need greater accountability and openness.

Heroes, and others

In such circumstances, self-promoters reign. Researchers come up with sometimes outlandish projects, and sell them to bureaucrats and politicians. The politicians use scientists to further their own ambitions, hailing the researchers as new heroes of Chinese science and giving them ever-bigger budgets. Chinese researchers in Beijing and Shanghai, as well as those overseas, condemn this "insiders' game".

The recent funding boom included several programmes aimed at bringing scientists home from abroad; there is no shortage of Chinese science success stories, especially in the United States. But although many of these returnees are China's most qualified and experienced researchers, there are also many — known in the overseas Chinese community as "scientific scammers" — who overstate their experience and expertise in order to gain the ear of the right government official. With that done, all is settled: a new hero of Chinese science emerges.

Once a project starts, it is difficult to criticize. Failed projects continue because the government will not acknowledge their failure — it does not want to lose face. A few years back, "nucleic acid biscuits" became a hot item, based on the logic that nucleic acids were essential nutrients. This was supported by some academic reports, and the biscuits were sold with government blessing, while criticism from those who doubted that DNA made good food was ignored or hushed up. As one overseas Chinese said: "fake drugs based on fake papers."

There is unlikely to be much benefit from the new rewards for productivity from the Knowledge Innovation Program, which began in 1998. In the 1990s, publication in Western journals was encouraged, reversing a past policy by which Chinese researchers were obliged to publish in Chinese journals. But the current policy goes too far. Monetary rewards are given for publications, and their value is based on the level of the journal in which a paper is published. Many researchers fear the effect of this short-term pressure to produce results. One member of the Chinese Academy of Sciences said he felt he was being micro-managed. Some ideas cannot develop in such a short-term fashion, and those with ideas requiring longer to come to fruition will be unable to get the required funding.

What is worse, the pressure to publish, for money or for status, along with an inadequate (or corrupt?) screening system, has led to the problem of plagiarism. This has even spread to the upper echelons of academia, tainting the name of a top executive at a company associated with Peking University.

Finding solutions

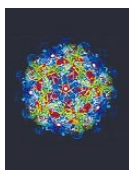
China is taking dramatic steps to catch up with the West. It supports many start-up companies through university spin-offs. But often the university is a major shareholder, and there is justified scepticism about the profitability of such companies. More probably, they will continue to depend on government support.

In the absence of an appropriate base for decisions, new initiatives with apparently exciting potential will lead to failed projects, loss of confidence and further obstacles. The National Natural Science Foundation of China and the Chinese Academy of Sciences are making some moves to improve the peer-review system. But even with attempts to encourage peer review of anonymous papers, the community is so small that reviewers can easily tell who the applicant is. A larger community of researchers needs to be tapped as reviewers. Possible candidates are Chinese people in the United States (and in Taiwan, in an ideal world). And if proposals were prepared in English, they could have an even larger pool of reviewers.

The Chinese scientific community has great scientific potential. But the government risks alienating the very people who have this potential. Researchers seem willing to leave good jobs in the United States — for the good of China and Chinese science. But their good intentions are likely to be dashed if the government is too closely involved in too many areas. One returning scientist was cautioned by his Chinese mentor in the United States: "It's great to go back, but just don't become an official."

The latest investment package from the Ministry of Science and Technology is aimed at original research rather than following up on high-technology trends. This attempt at self-reliance is understandable. But cross a thin line, and you get what many Chinese themselves would say is an essential Chinese trait — the desire to do everything oneself. Overcentralization and a concentration of decision-making could greatly inhibit progress. China needs to open up its science policy, to delegate more of its science management, and to establish and maintain more channels of communication inside and beyond its borders.





Foot and mouth
Effective vaccine
treatment for animal
virus remains elusive
p4



Lab shake-up
Drugs giant breaks
up research into
autonomous units
p5



Genetic rift
Mendel's heirs argue
over how he might be
best remembered
p6



Value added
New goal-oriented
research must prove
its worth in Japan
p7

Bush declines to support drug companies' line on AIDS profits

David Dickson

The administration of president George W. Bush has surprised international aid organizations and AIDS activists by indicating that it will take a liberal attitude towards developing countries' efforts to produce generic versions of expensive drugs.

The move coincides with a heated public controversy over the situation faced by sufferers of diseases such as AIDS, who are unable to afford even discounted prices.

Next week, for example, a South African court will hear a challenge by 42 pharmaceutical companies against the government for allowing the domestic production of cheaper versions of their anti-AIDS drugs.

Meanwhile Father Angelo d'Agostino, head of Nyumbani Orphanage for children with HIV or AIDS in the Kenyan capital Nairobi, said last Friday that he plans to accept an offer by the Indian company Cipla of Mumbai, which has promised to make the drugs available at about one-third of the minimum cost on offer from the Western companies (see *Nature* 409, 751; 2001).

The previous US administration decided in 1999 to take a more relaxed attitude to developing countries wishing to produce their own versions of expensive anti-AIDS drugs patented in the West. Bush had been widely expected to reverse this decision.

But last week, Robert Zoellick, the US trade representative, said that the United States was not planning to change the current policy. He promised to back efforts by the administration "to work with countries that develop serious programmes to prevent and treat this horrible disease", provided such efforts were "consistent with our overall efforts to protect America's investment in intellectual property".

The statement's significance was played down by the Pharmaceutical Research and Manufacturers of America, which has been pressing for the United States to take action against countries such as Brazil which it claims are undermining US patent interests by producing generic drugs.

But it was warmly welcomed by AIDS activists and public-interest groups. They are hoping that the embarrassment being



Hopeful sign: Christine Gatwiri, 7, at Nyumbani.

caused to major drug companies such as Boehringer Ingelheim and GlaxoSmithKline will persuade them to modify their stance.

Some are saying the publicly funded laboratories which carried out research leading to the development of these drugs should also become involved. James Love, director of the Washington-based Consumer Project on Technology, says US institutions that own patents on important HIV drugs should endorse Cipla's request for a licence to sell its

drugs in developing countries where the Western patents are valid, paying the patent-holders royalties of up to 5%.

"Where licences cannot be obtained voluntarily [for the domestic production of generic anti-AIDS drugs], it is now time for the World Health Organization, UNAIDS, the World Bank, governments or non-governmental organizations to obtain compulsory licences," he adds. "It is morally repugnant to delay this process any longer."

But others are arguing that the situation should be addressed by increasing the amount of money available through aid programmes to purchase anti-AIDS drugs.

Meanwhile the aid agency Médecins sans Frontières (MSF) has agreed to distribute Cipla's three-drug antiretroviral cocktail free of charge in 10 countries. "Wherever it is purchased by governments, it would be sold at reduced prices," Daniel Berman of MSF said last week. He said the agency would integrate some Cipla drugs into antiretroviral pilot programmes in 10 countries including Thailand and Cameroon, where the programme has already started. ■

Growers cotton on to GM bollworm

Rex Dalton, San Diego

The first US field test of a genetically modified insect is being planned in Arizona. Organizers say its outcome could help control the pink bollworm, a cotton pest.

Researchers from the US Department of Agriculture (USDA) and the University of California (UC) at Riverside will issue a public notice in a few weeks, kicking off the process needed to obtain permission for the test.

About 3,600 moths, which produce the bollworm larvae that burrow into cotton, will be studied in a one-hectare enclosure in a field near Phoenix. They will have been modified with a genetic marker — enhanced green fluorescent protein, from a jellyfish — so that they can be tracked.

Later, the team plans to introduce a 'lethal gene' designed to kill the progeny of

both engineered and naturally occurring moths. Testing of moths with the lethal gene will require an additional review period and permit, federal officials say.

The method was developed by a team headed by UC Riverside entomologist Thomas Miller and USDA entomologist Robert Staten in Phoenix, with a \$1 million grant from cotton growers. They say that the moths are an invasive species that plays no role in the local ecosystem.

But Charles Margulis, who monitors genetic issues for Greenpeace, calls the tests "a clear and present danger of irreversible damage".

John Caravetta, of the Arizona Agriculture Department, which has veto power, said the state supports the concept, but needs to see the final plan before commenting in detail. ■

Farmers act to avert foot-and-mouth crisis

Jim Giles, London

Beleagued British farmers are reeling from another animal health crisis this week, after cases of foot-and-mouth disease were confirmed at locations across the country. It is the first major outbreak of the highly infectious disease to affect Britain in over 30 years.

The government acted quickly to halt movement of animals after vets from the government-funded Institute for Animal Health (IAH) found cases at an abattoir in Essex.

Eleven further outbreaks had been confirmed by 27 February, including one at a sheep-exporting business in southwest England. The possibility that the disease may have been exported led France, Germany, Belgium and the Netherlands to introduce measures restricting movement of animals that had recently arrived from Britain.

Intensive farming practices and the increased movement of livestock, which has resulted from the closure of local abattoirs in response to the outbreak of bovine spongiform encephalopathy, have been blamed in the press for the scale of the outbreak. But experts say that these factors are of little importance.



Dead meat: pyres burn in Britain as farmers struggle to contain foot-and-mouth disease.

Peter Roeder, an animal health officer at the United Nations' Food and Agriculture Organization, points out that foot-and-mouth disease is already endemic in some areas of southern Africa and parts of Asia. Imports of foodstuffs from these regions are strictly controlled, but customs officials say it is impossible to block illegal imports completely.

The UK virus is one of seven known types of the disease. Called type O, it originated in India before spreading east into China, Japan, Korea and Taiwan. It arrived in South Africa last year, but has been eradicated there by slaughtering potentially infected cattle.

Slaughtering of livestock has already begun in Britain, as existing vaccines can stop symptoms but do not prevent animals from carrying or spreading the disease. "The virus can exist in vaccinated cattle for up to three years," says the IAH's Paul Kitching, "and trade restrictions reflect this. To resume trade a country has to be totally free of the disease."

Making the vaccine, which requires large amounts of the live virus, is also risky. Outbreaks from vaccine production centres have persuaded several European countries to end vaccination programmes.

Vaccine development is further complicated by the virus's high mutation rate. This leads to the existence of numerous subtypes within each type, some of which may require different vaccines. The IAH is responsible for matching vaccines to the hundreds of samples it receives from around the world every year. It dealt with 400 samples last year but, says Kitching, the true number of outbreaks is "uncountable".

The IAH, Pfizer and Eli Lilly each have groups working on new vaccines. "Safer, more stable vaccines are possible," says Kitching, "but it is highly unlikely that a new vaccine will eliminate all the problems." ■

Europe frames fresh funding initiative for research

Quirin Schiermeier, Munich

Sweeping reforms to the Framework programme, the European Union's main research support activity, have been proposed by the European Commission.

Instead of supporting large numbers of unrelated projects, Framework's next phase will fund a much smaller number of large, integrated projects. The commission hopes that the change will distinguish Framework from the plethora of national research funding agencies in Europe, and make a real difference to European competence in key scientific areas.

The new programme means a particularly drastic change in the life sciences, where four-fifths of existing projects will lose support, as the focus is shifted to functional genomics.

The commission is proposing a budget of 17.5 billion euros (US\$16 billion) over five years for the next — sixth — Framework programme, which starts in 2003. This is a 17% increase on the current fifth Framework programme's budget.

But the average size of Framework grants — currently around 1.5 million euros per project — will rise by an order of magnitude. Proposed 'networks of



Philippe Busquin: chasing his vision of integrated research in Europe.

excellence' in critical scientific disciplines will, for example, receive tens of millions of euros each year.

To attract Europe's best scientists, the commission has also suggested ways of simplifying grant application procedures and project management.

The reforms are aimed at improving interaction between national and European programmes, officials say. According to one official, the commission plans to operate in

"catalytic mode, building on, rather than duplicating, what happens in the member states".

The new networks of excellence are designed to complement the concept of a single 'European research area' proposed by Philippe Busquin, the European Commissioner for Research. This aims to make the best use of Europe's research resources (see *Nature* 405, 873; 2000).

The plan will also back large-scale integrated projects designed around specific technologies or research applications — for example, high-performance materials or information technology. Ideally, these will be conducted as partnerships between academia and industry.

Grant recipients will be allowed to take on partners as their project proceeds, instead of having them all in place at the start. And in some areas grant applications will be split into two phases, with a preliminary application being used to identify a project's most promising ideas.

The proposal is being sent to the European Union's member states for consideration. A decision on its adoption will be made next year through the Council of Ministers and the European Parliament. ■

GlaxoSmithKline pushes its labs towards 'biotech' future

David Adam, London

One of the world's largest drug corporations, GlaxoSmithKline, has announced radical plans to rearrange its research laboratories into six units that, it says, will each walk, talk and look like individual biotechnology businesses.

But some of the 16,000 researchers at the company, which was founded last December by the merger of GlaxoWellcome and SmithKline Beecham, are already expressing doubts that the shake-up will have the desired effect. They are dubious about how much autonomy the research units will be allowed, and warn about the impact on staff morale of post-merger jockeying for position.

GlaxoSmithKline says that the new units will operate as internal, competing biotech businesses. Its chief executive, Jean-Pierre Garnier, argues that the units are the best way to turn ideas into new drugs. "To take a target and transform it into a product that works in the clinic, you have to be nimble," he says.

The autonomous units will take validated drug targets, investigate dose and delivery methods, and then test how safe and effective they are in animals. The units will also help to carry out the initial stages of human tests but will then hand over responsibility to the core company for large-scale, expensive clinical trials.

Garnier adds that the earlier stages in the drug-discovery process, including research in functional genomics and high-throughput screening of therapy candidates, will continue to be a large operation, integrated across several of the newly merged company's sites. "We start big, we move to small and then back to big again," he says. "Nobody has ever tried to do this before."

Three of the new drug-discovery centres, which will compete among themselves for company resources, will be located in the United States, two in Britain and one in Italy. The company's laboratory in Upper Providence, Pennsylvania, will work on anti-infective medicines; Upper Merion, Pennsylvania will tackle cardiovascular, oncology and genitourinary conditions; and research into metabolic control and antivirals will be based at Research Triangle Park, North Carolina.

In Europe, researchers in Verona will develop psychiatry drugs leads; neurology drug candidates will be investigated at Harlow; and the Stevenage site will handle anti-inflammatory research. There will be a separate division for vaccines.

"The people in the centres will be rewarded strictly on what they produce and the scientists behind them will be very account-



Rapid response: GlaxoSmithKline's Harlow laboratory is gearing up for autonomy.

able. They can't hide and will have to produce drugs," Garnier insists.

The move to create the units follows the trial isolation of SmithKline Beecham's antibiotic and antiviral research activities. "It helped people to move faster and now we want to do it on a much bigger scale," Garnier says. "We are pretty certain it's going to work."

Tadataka Yamada, head of research at GlaxoSmithKline, says the changes will introduce a range of incentives to mimic the workings of biotech companies, including stock options, royalties on drug sales, and bonuses for both groups and individuals. Yamada also promises "high-profile recognition within the company" for existing and emerging stars.

But the company has already announced that two research centres will close and more could follow, and some GlaxoSmithKline researchers are expressing concerns about the integration plans. Several employees said privately that they are uncertain about how well the units will work. They warn that the large gravitational pull of the central corporation will continue to dominate the trajectories of the smaller research units orbiting it.

Some of the researchers also noted that a previously successful Glaxo drug-discovery unit, based on a similarly entrepreneurial model and located within the pharmacology department at the University of Cambridge, is being closed by the merged corporation. ■

http://www.gsk.com

Hepatitis pioneer takes the reins for French medicine

Declan Butler, Paris

Christian Bréchet, a clinical researcher specializing in liver disease and hepatitis, has been named as the new director of INSERM, France's main biomedical research agency.

Bréchet is best known for his work on hepatitis B and C — especially his investigation of the role of viral hepatitis in the development of liver cancer (see *Nature* 286, 533–535; 1980).

After his appointment, Bréchet said in a newspaper interview that one of his priorities would be to strike the right balance between clinical and basic research at INSERM. He also plans to develop informal research groups within the agency to act as special "strike forces" in important areas of multidisciplinary research. And he says he will rejuvenate INSERM from the bottom up, by supporting talented young researchers.

Bréchet's appointment was announced last week by Roger-Gérard Schwartzberg, the French research minister. Schwartzberg says he wants Bréchet to give priority to emerging areas of research such as gene therapy, and to areas of new relevance to public health, such as the study of prion-related diseases.

The research minister also wants to see better coordination between INSERM's 256 research units, and more cooperation with other research sponsors in France.

Bréchet's dual role as a practising clinician, at the Necker Children's Hospital in Paris, and as a researcher — he is director of an INSERM unit at the Pasteur Institute — apparently contributed to his selection by Schwartzberg. "It is important to have someone who has direct and regular contact with patients," says the minister. ■



Christian Bréchet, INSERM's new chief.

Benefits of low-tar smokes just a pipe dream

Corie Lok, Washington

Smokers puffing on low-tar cigarettes may be deluding themselves, according to the US Institute of Medicine (IoM). Research has yet to show that these modified cigarettes are any safer than their conventional counterparts, a report by the IoM concludes.

In its study, the IoM found no evidence that using low-tar cigarettes reduces a smoker's exposure to carcinogens. And even if these cigarettes do lower the amount of carcinogens inhaled, there is also no evidence that this would necessarily reduce the smoker's susceptibility to disease.

"No one knows whether decreased exposure translates into decreased risk for disease," says Stuart Bondurant, medical professor at the University of North Carolina, Chapel Hill, and chair of the IoM committee that studied the problem.

The independent panel made its assessment of the harm reduction potential of low-tar cigarettes and other nicotine products at the request of the US Food and Drug Administration (FDA). Its report, "Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction", was published last week.

Cigarettes that are low in tar and nicotine raise a public health concern because they may use implied but unsubstantiated safety claims to keep smokers hooked on tobacco and attract new users, the report says. And

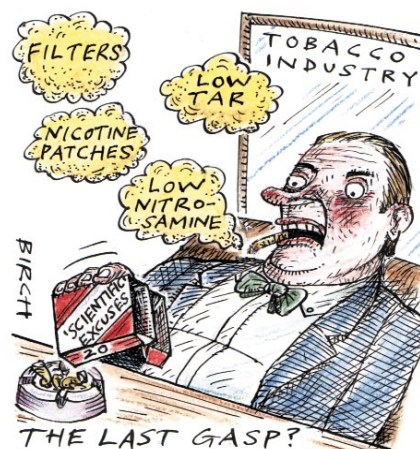
there is no guarantee that the products reduce harmful exposure because the smokers tend to suck harder and smoke more frequently to maintain their nicotine levels.

The IoM panel found that more research is needed to establish whether there is a link between exposure reduction and harm reduction. Such research, it says, could improve understanding of dose-response relationships of specific smoke components, develop and validate biomarkers that link smoke exposure to disease risk, and explore the epidemiology of smoking and health.

The report also suggests preclinical studies to determine harmful smoke components and their pathogenic effects. For example, new genomic and proteomic technologies could be used to look at the effect of tobacco exposure on gene translation and expression in disease.

FDA-approved nicotine products were also examined by the panel. These include patches, sprays and chewing gum, which deliver pure nicotine at low doses. The products are approved only for short-term use to help smokers quit the habit, but research has not established long-term safety for those smokers using them to continue to get low levels of nicotine, the report says.

The report recommends new regulations requiring manufacturers to disclose ingredient, exposure and experimental data to back



up any risk-reduction claims. But it did not suggest a particular agency to take on this regulatory task. It also emphasizes the need for a national, comprehensive surveillance system to monitor product characteristics and usage patterns.

Although both the National Institutes of Health and the tobacco industry have conducted some research into the health effects of the products covered in the report, the field is "in its infancy," according to Peter Shields, a panel member and associate director of Georgetown University's Lombardi Cancer Center.

♦ <http://books.nap.edu/catalog/10029.html>

Museum suffers spiritual cramps over Mendel's work

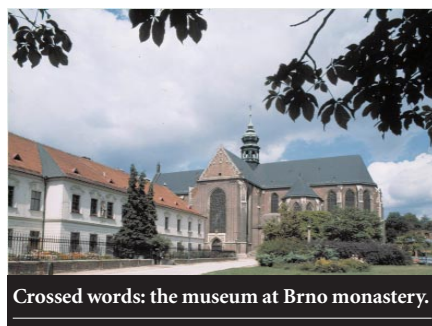
David Adam

With the ink on the published human genome barely dry, an unseemly row has broken out at the monastery where Gregor Mendel founded the study of genetics almost 150 years ago. The argument is over the commemoration of his life and work.

Mendel first entered the monastery at Brno in the Czech Republic in 1843, and was elected abbot in 1868. His study there of how pea plant characteristics passed between generations revealed the fundamental laws of genetic inheritance.

Now staff at a small museum that is situated in the monastery say that the current abbot, Lukas Evzen Martinec, is threatening to evict them because their museum places too much emphasis on the scientific aspects of Mendel's life. The abbot, they say, wants a new exhibition of Mendel's life that will reflect his religious beliefs as well as his scientific interests.

Martinec confirms that he is unhappy with the low religious content of the museum's display. But he says that his decision to replace it is part of wider plans to



develop a new, extended museum. He adds that he gave the existing museum's directors more than a year to come up with a proposal for a new exhibition, and that, having failed to do so, they are mounting a "disinformation campaign" against him.

"Communication with the abbot is nonexistent, but we never believed he would go this far," says Anna Matalova, director of the museum, the Mendelianum.

The museum's display was closed earlier this year and the exhibition hall is due to be refurbished. "The abbot refused to promise that the Mendelianum would be returned to

the exhibition hall after the reconstruction," says Matalova.

The museum also has archives and offices in the monastery, but Martinec recommended in a recent letter that they "consider immediate relocation". The state-run museum rents the rooms from the monastery, but pays less than a private company would.

Martinec says that the exhibition hall has been "neglected" after "35 years without adequate maintenance". He is also angry that the museum's owners have not uncovered religious motifs that were hidden during the communist days. He is now preparing a replacement exhibition with help from local university researchers.

Emil Palecek, a molecular biologist at the Czech science academy's Institute of Biophysics in Brno, has seen the abbot's plans and predicts that the new exhibition will be an improvement. He says the refurbishment is an important part of a plan to transform that part of the monastery into a conference centre, and perhaps even a bioinformatics institute.

Goal-directed revamp for Japanese research

David Cyranoski, Tokyo

An ambitious effort to restructure the research arm of Japan's industry ministry has already attracted a clutch of prestigious scientists to head its new institutes. They include leading university researchers from the United States and Sumio Iijima, the discoverer of carbon nanotubes.

Hiroyuki Yoshikawa, a former president of Tokyo University, is in charge of the research revamp at the Ministry of Economy, Trade and Industry (METI), previously known as the Ministry of International Trade and Industry (MITI).

On 1 April, the National Institute of Advanced Industrial Science and Technology (AIST) will be set up as an autonomous agency within METI. It will replace the old Agency of Industrial Science and Technology (also known as AIST).

The new body will comprise 45 institutes and centres. Its structure is designed to combine application-orientated goals with considerable operational autonomy for the institutes.

Yoshikawa will appoint institute directors and allocate funds, powers previously held by civil servants at METI. But Yoshikawa says he recognizes that autonomy brings with it the responsibility for delivering results. "A national institute has to prove its value to society — it has to be mission-oriented," he says. He adds that the old AIST allowed researchers too much freedom "to follow their own curiosity".

One key mission will be to gear research towards "sustainable development". The aim is not only for advances in environmental science, but also to achieve social goals, Yoshikawa says. Japan's rapidly ageing population, for example, might benefit from the applications of robotic technology.

The new institutes and centres will be formed by splitting up the old AIST's 15 laboratories. Each institute will house fewer than 100 researchers, whereas the old AIST laboratories employed as many as 600 research staff on a single site. The reduced size is meant to bring greater accountability, and the new system will be evaluated in three to five years' time.

Although researchers appreciated their freedom, under the old system, from having to consider practical applications for their research, some say they are ready for tighter organization. "To make a big impact, you need strong management and clearer authority," says Junji Itoh, a 16-year veteran of the old AIST and director-designate of the new Electronics Research Institute, which, like most of the institutes, is based in Tsukuba.

But some critics fear that too many institutes will lead to inefficiency. "Communication will be key," says Fu-Kuo Chang, a



Fresh horizons: the AIST facilities at Tsukuba, which will be led by Hiroyuki Yoshikawa (inset).

researcher in aeronautics from Stanford University in California, who will head the Smart Structures Research Center.

The institutes are being designed to forge stronger ties between basic science and its industrial application. Previously there have been few such links in Japan. New measures to give researchers a direct share in any technology transfer agreements are designed to encourage such collaboration. The institutes will also accept commissioned research projects from commercial companies.

Iijima will head AIST's Research Center of Advanced Carbon Materials. And Kotoku Kurachi, a professor of human genetics, is

returning to Japan after 30 years in the United States to head the Gene Discovery Research Center. He says the emphasis on practical applications "is a new thing in Japan, where academics have generally stuck to a 'pure science' ethic". Kurachi, who was most recently at the University of Michigan, says he plans to work on medical applications of his research on the genes involved in the increased risk of cardiovascular disease with age.

The transformation of AIST is being watched with special interest by Japan's universities, which are also due to become independent agencies in 2003 (see *Nature* 401, 416; 1999).

Israel seeks sweet smell of success

Haim Watzman, Jerusalem

A project to create a DNA database of the genes responsible for the perfume in flowers is under way at the Hebrew University of Jerusalem. Its purpose is to identify genes that can be inserted into cultivated flowers in order to improve their smell.

The project has already identified some 2,000 genes that are active in rose petals and has created a library of expressed sequence tags (ESTs) that map the genes.

According to David Weiss, a horticulturist and a member of the project team, the work has identified seven candidate genes for initiators of the biochemical processes that produce rose aromas. The next step, he says, is to study the biochemical pathways that lead from the gene, through the proteins they produce, to the substances that give the flower its pleasant smell.

"We know a lot about the chemistry of flower perfumes, many of which are produced synthetically by industry," Weiss

explains. "But we know very little about the biochemistry — the biological processes that produce the perfumes." The work is supported by a US\$220,000 annual grant from the Ministry of Science.

Weiss and his colleague Alexander Vainstein recently broke new ground. They transferred an aroma-producing gene isolated by Eran Pichersky of the University of Michigan from a wild California flower, the clarkia (*Clarkia breweri*), into carnations. Carnations possessing the gene, Weiss says, have a new biochemical pathway that produces the major compound responsible for the clarkia's aroma, linalool, and other fragrance-producing chemicals.

Cultivated flowers, bred for their large, colourful blooms, tend to lose much of the smell of their wild ancestors, Weiss explains. Getting the smell back into them could be a major boon for Israel's 1,600 flower-growers, who export \$250 million worth of flowers each year.

European biology service gets public backing

Paris The European Commission has approved a grant of 2.4 million euros (US\$2.2 million) over three years to help develop a Europe-based global biological information service.

E-Biosci, as the service is to be called, will be a European counterpart to the PubMed service run by the US National Center for Biotechnology Information. It will link literature, gene and protein databases and other biological resources, and is scheduled to open in July.

The service was initially proposed by the European Molecular Biology Organization (EMBO) in response to US plans to create PubMed Central, a free website for life-science papers (see *Nature* **401**, 413; 1999).

NSF to support botanists in coordinating genome work

San Francisco The National Science Foundation (NSF) has said it will support a wide network of botanists in efforts to coordinate future plant genome research.

The network, announced at last month's annual meeting of the American Association

for the Advancement of Science, will be charged with agreeing on the most important plants to target for genome sequencing. Its work should also help the NSF to select research projects in plant genetics and evolution, participants say.

The NSF will grant \$500,000 to the network, which will be called Deep Gene, over five years. It also announced a companion project, called Deep Time, to examine fossil records and correlate them with discoveries from plant genetics.

Animal-lab director injured in attack

London The managing director of a UK research laboratory targeted by animal-rights protesters was attacked last week by three masked assailants wielding baseball bats.

Brian Cass, head of Huntingdon Life Sciences (HLS), was attacked as he arrived at his home in St Ives, Cambridgeshire. Police say they have no doubt the attack was linked to his involvement with animal experiments.

Cass was beaten about the head and upper body and needed several stitches in a head wound. A passer-by who tried to help him was sprayed in the face with CS gas. Animal rights groups, including Stop Huntingdon Animal Cruelty which has led the HLS protests, condemned the attack.

'Peer miles' could make reviewers fly through work

London Punctual reviewing of grant proposals will earn researchers points and their departments prizes, under a new incentive scheme to be launched in Britain. Already dubbed 'peer miles' by some, the scheme will award cash to the departments of researchers who deliver assessment reports on time, suggest alternative reviewers and attend meetings.

The Engineering and Physical Sciences Research Council (EPSRC), a government-funded grant-awarding body, expects to give away about £750,000 (US\$1 million) to its reviewers during the scheme's one-year trial.

"The sums involved aren't vast, but big departments could perhaps get up to £20,000," says David Clark, director of research and innovation at the EPSRC, who admits that the idea is "borrowed from air miles and supermarket loyalty schemes".

Nobel laureates back stem-cell research

Washington Eighty Nobel laureates wrote to US President George W. Bush last week asking him to continue federal support for embryonic stem-cell research. Researchers have until 15 March to apply for funds for

stem-cell research, but fear that Bush may ban the research before then.

The US National Institutes of Health last year approved guidelines for government funding of stem-cell research, but not their derivation from human embryos. Instead the cells have to be extracted from 'left over' embryos created during *in vitro* fertilization. Anti-abortion groups have lobbied against the work, and the Bush administration has hinted that it will reverse the decision.

The letter echoes the stance of patient groups who back embryonic stem-cell research because of its potential to treat disease.

Centres to study threat of illegal aliens in the water

Washington Two major US ecological research centres are launching a new collaboration to study the threat of non-indigenous aquatic species in US waters.

The Smithsonian Environmental Research Center in Edgewater, Maryland, and the Florida Caribbean Science Center run by the US Geological Survey (USGS) are teaming up to research non-native plants, animals and microorganisms in US ecosystems. The USGS believes invasive species are the second biggest threat to native biodiversity, after loss of habitat.

The collaboration will make use of an existing USGS database on invasive species in freshwater species and the Smithsonian's data on non-native organisms brought into US waters by ballast water used on ships.

Mouse cancer model repository opens

Washington The US National Cancer Institute's mouse cancer model repository has declared itself open for business.

Located at the Cancer Research Center in Frederick, Maryland, the repository has breeder pairs of two strains available, and plans to add 30 more strains by the end of the year. Researchers have been invited to submit strains and suggest desirable ones. Samples of less popular strains will be cryopreserved.

► <http://web.ncifcrf.gov/researchresources/mmhcc/default.asp>

Very short flight of NASA's long-duration balloon

Sydney A NASA project to study cosmic radiation is on hold after a giant data-gathering balloon sprang a leak just hours after being launched on its maiden flight last weekend. The inappropriately named Ultra Long Duration Balloon was brought down by remote control. NASA scientists are now



Grounded: what burst NASA's big balloon?

trying to figure out what went wrong.

The 60-metre-wide balloon, which is seen as a potentially cheap alternative to satellites, should have circumnavigated the globe for two weeks, skimming along the top of the Earth's atmosphere. NASA is hoping to launch a replacement balloon within a week.

Correction: The News item "BSE crisis sinks German public biotech programme" (*Nature* 409, 549; 2001) quoted Andreas Tierfelder, spokesman for the German subsidiary of Monsanto, as saying that the cancellation of the programme "is a substantial loss of image to our business". *Nature* accepts that Monsanto does not view its image to have been damaged in any way. Monsanto intends to participate in the programme, if it continues at a later date.

A great leap forward

After helping to sequence the human genome, Chinese scientists are debating how best to continue the push towards becoming a world power in biology. David Cyranoski reports.

China wears its participation in the international Human Genome Project (HGP) like a badge of honour. Last month, when the HGP published its sequence of the human genome in *Nature*¹, Chinese institutes were among those sharing the spotlight.

"Now everyone in China knows what the human genome is," says Huanming Yang, director of the Beijing Genomics Institute (BGI). "Chinese people are so proud to be part of this international effort."

In 1998, China's scientific leaders overcame scepticism from some members of the HGP — and from many of their own researchers — to become the only developing country to take a role in sequencing the human genome. The Ministry of Science and Technology launched its Chinese National Human Genome Center (CHGC), with branches in Beijing and Shanghai. A year later the Chinese Academy of Sciences opened the BGI. These institutes contributed 1% of the published sequence — an achievement that is of huge symbolic importance.

Having now proven its mettle in sequencing, and following earlier successes in plant transgenics, China is now taking a leap towards assembling all the elements needed

to become a major force in biology and biotechnology. With strong government support, researchers are setting up programmes in everything from stem-cell research, through large-scale efforts to determine protein structures, to population studies to hunt for human disease genes. There is even talk of trying to clone the endangered giant panda.

For researchers returning from positions abroad, the pace of change is striking. Although most of China's biologists still work in run-down buildings, these labs are starting to bulge with gleaming new equipment. And many institutes are planning to expand their facilities — just part of the construction upheaval that is currently transforming China's cityscapes. Returning after three years at the Memorial Sloan-Kettering Cancer Center in New York to Shanghai, a city he "no longer recognized", molecular geneticist Zhu-Gang Wang of Shanghai Second Medical University was startled to find that reagents could now be acquired overnight, rather than in the months that he remembered.

But amid this rapid growth, some researchers are warning that there is more to be done than simply matching the spending of the established powers. As Chinese biologists

debate the optimal balance between building more capacity for DNA sequencing against investing in areas such as functional genomics and medical genetics, there are calls for more coordination, and better peer review to judge the relative merits of the various programmes being considered. Other researchers are concerned that, despite moves towards a market economy, China still lacks the genuinely competitive environment needed to nurture a vibrant biotechnology industry.

A valuable lesson

Although many Chinese researchers originally questioned the value of a Chinese contribution to the HGP, there is now little doubt that the work has provided a springboard for the development of Chinese biology. "We learned large-scale sequencing, how to coordinate many centres, and how to do bioinformatics," says Zhu Chen, director of the Shanghai CHGC and vice-president of the Chinese Academy of Sciences.

Now, those who led China's contribution to the HGP are debating how to strike a balance between building on these strengths and exploring other frontiers. At the BGI, Yang is pressing on with sequencing. In late January, he opened a new sequencing facility in Hangzhou, south of Shanghai. This will work on two main complete genome sequencing projects: the pig — a collaboration with Danish scientists — and the parent strains of China's 'superhybrid' rice, a productive variety that accounts for 60% of the country's rice harvest.

"Some people say, 'sequencing is not everything,'" says Yang, "but I believe that it is the best place to start." However, Yang feels he is losing a tug-of-war over central government funding to those who are switching to functional genomics, research that aims to



The Beijing Genomics Institute made an important contribution to the human genome sequence.



Food for thought: Chinese researchers are turning their attention to the genome of 'superhybrid' rice.

identify the biological role played by the newly sequenced genes. To open the Hangzhou facility, Yang had to borrow US\$6 million from the city government.

Although Chen's Shanghai CHGC was responsible for about a quarter of China's HGP contribution, he is now more excited by the possibilities offered by the hundreds of full-length complementary DNAs sequenced as part of China's involvement in the project. These represent the sequences of expressed genes, and can be engineered into cells which are then cultured to mass-produce proteins so that their structures can be determined.

Like the Shanghai CHGC, the Beijing branch is turning increasingly to functional studies. The two centres have several joint projects, focusing on genes related to liver cancer, nasopharyngeal cancer, oesophageal cancer and leukaemia.

But there is one major point of convergence in the research strategies of the BGI and the CHGC: both are using their sequencing machines to identify single-nucleotide polymorphisms (SNPs) from Chinese populations. These markers of genetic variability should aid the search for the genes that predispose humans to complex diseases such as cancer and psychiatric disorders. Some critics argue that the Chinese SNP projects will largely duplicate the efforts of the international SNP Consortium, a collaborative effort uniting leading Western genome centres and drugs companies. But Chen argues that about 30% of the SNPs in Chinese populations will be different from those found elsewhere — twice as many as has previously been thought.

Whether or not a separate Chinese SNP effort is required, there is widespread agree-

ment that China has a lot to offer geneticists. The country's array of isolated, relatively homogeneous sub-populations from 56 different ethnic groups represents a unique resource for disease gene searches. Inbred-

ing has produced a homogeneous genetic make-up within each population, and the lack of emigration or immigration means that it is easy to construct large family pedigrees. These features have made Iceland and Finland hotspots for mapping the chromosomal location of genes predisposing to disease. "We have several Icelanders and several Finns," says Lin He of Shanghai Jiao Tong University, who coordinates a bank for blood, tumour and cerebrospinal fluid samples collected from populations across China.

The search for genes that underlie diseases, and for ways to block their effects, is a national priority. "With an estimated population of 1.6 billion by 2030 and an ageing society brought on by the one-child policy, we will need medical breakthroughs," says Xu Zhi-Hong, president of Peking University. "This is our challenge."

It is a challenge that both the government and researchers are taking on. In 1998, the government set up a programme alongside the HGP sequencing effort, which has established banks of blood samples and associated information about family pedigrees in Hunan, Guanzhuo, Beijing and Shanghai. Independent scientists have been working in parallel. Yunnan University recently announced that it has assembled the largest biological sample collection in the world in terms of ethnic diversity. At Shanghai Jiao

Politics, ethics and collaborations

As China's medical geneticists move to exploit the unique resource offered by the country's isolated, genetically homogeneous populations (see main text), there should be opportunities for fruitful collaborations with researchers from abroad. But such collaborations are still difficult. On the Chinese side, geneticists are concerned at becoming little more than 'sample vendors' for their foreign partners. Many Western geneticists, meanwhile, remain uneasy about China's flirtation with eugenics.

In 1997, China's Premier, Jiang Zemin, threw medical genetics into the spotlight by commenting publicly on the country's genetic heritage. "One without foresight will encounter sudden doom," he said. "We must treasure our genetic resources." At the time, several projects had raised concerns about whether China's genetic resources were being plundered by visiting scientists. This led to formation of the Human Genetic Resources Administration

of China, which must now approve any research using human samples for genetic studies.

The administration's guidelines state that, among other conditions, both sides of any international collaboration will share in resultant intellectual property. But opinions are divided on whether the potential for abuse has been removed. "The guidelines are clear, and 90% of international projects are approved," says Zhu Chen, director of the Chinese National Human Genome Center in Shanghai. But Huanming Yang, director of the Beijing Genomics Institute, is unconvinced. "Regulations are still too loose," he says.

Geneticists visiting China sometimes have doubts about the quality of their collaborators' family pedigrees and samples. "People will say they're related to anyone if they think they can get free medical treatment or some money by participating," says one Chinese biotechnology veteran. Some Western researchers also question

the accuracy of clinical diagnoses. "I would guess schizophrenia diagnosis in a totalitarian country would be far from reliable," says one US-based geneticist.

The political climate in China also raises deeper ethical concerns. Many geneticists were appalled when, in 1995, the Chinese government introduced the 'Mother and Infant Health Care Law'. This required couples planning marriage to undergo genetic screening, and included provisions to encourage sterilization of those found to be carrying disease genes.

In protest, many geneticists threatened to boycott the International Congress of Genetics, held in Beijing in 1998. Although the controversy was partly defused by a government promise to rewrite the law, the changes have yet to be made. Last December, the Ethical, Legal and Social Issues Committee of the Chinese Human Genome Project renewed its call for the promise to be acted upon.

news feature

► Tong University, meanwhile, He claims to have compiled the world's largest sample bank related to neuropsychiatric disease, in a joint project with the Shanghai Institutes of Biological Sciences.

Rising stars

Studies of samples taken from Chinese populations are already yielding advances. The first success came in 1998, when a team led by Jia-hui Xia of the National Laboratory of Medical Genetics of China in Changsha, Hunan province, identified a gene for a form of neurological deafness². Then, last year, researchers led by He mapped the gene for brachydactyly type A-1 — a disease in which joints of the fingers are missing, misplaced or disfigured — to a small region of chromosome 2 (ref. 3). He and his colleagues say they have now isolated the gene, beating competition from several groups around the world, and have submitted the results for publication. "We just had good samples with good pedigrees," He says.

And the pace looks set to pick up. Last month, researchers led by Yan Shen of the Institute of Basic Medical Sciences in Beijing revealed the genetic basis of another disease, dentinogenesis imperfecta Shields type II, in which the teeth are discoloured and chip easily⁴. From their studies of large pedigrees in the Jiangsu province, north of Shanghai, the team showed that a mutation in a gene called *DSPP* is responsible. Alongside this paper, researchers led by Xiangyin Kong of the Shanghai Research Center of Biotechnology reported, from their studies of three extended families from the same region, that mutations in *DSPP* can also cause hearing loss⁵.

But there are obstacles to progress. Some Chinese people are unwilling to take part in studies for fear that discovery of a genetic disease could lead to discrimination. And in the longer term, China's one-child policy, and the increasing population mobility being promoted by the country's economic development, will pare down the number of large pedigrees and lead to genetic mixing between populations.

Commercial breaks?

To capitalize on the discovery of disease genes, biotechnology and pharmaceutical companies will have to take the lead. This is the weakest link in the chain, although with generous government support, universities are launching large numbers of spin-off companies. Peking University is the leader, boasting a portfolio of eight biotech firms.

"China's system for allowing professors to engage in industry is much more flexible than most others," says Jing Cheng, one of China's best-known biotech entrepreneurs, who runs a DNA microarray company spun off from Tsinghua University in Beijing. Indeed, compared with Pacific Rim neighbours such as Japan, there seems to be little or



Rich pickings: the relatively isolated sub-populations in China offer researchers a unique, homogeneous resource for investigating the genes underlying disease.

no restriction on university researchers trying their hand at business.

But unlike the situation in its capitalist neighbours, China's government retains a large chunk of any company, and may even fund its research. Some researchers argue that this is holding the sector back. "With generous government grants, people do not worry about making profits. Even if something is developed, it will sit on the shelf because no one will make the effort to commercialize," says one researcher working at an institute of the Chinese Academy of Sciences. "Many researchers seem just to use public funds freely to do research while claiming to set up their own company," says a senior executive of a Hong Kong-based biotech company.

Reliance on the government is "the greatest problem for biotech in China," agrees Yu-Min Mao, a geneticist at Fudan University in Shanghai. Venture capital is not well established, so would-be entrepreneurs have few other places to turn. But Mao is showing that it is possible to break the mould. In three short years, his United Gene Holdings has become China's biggest biotechnology company. It now covers a wide range of genomic science services such as gene cloning and sequencing, and claims to have identified 5,000 novel complementary DNAs.

Mao and his business partner, Yi Xie, invented their own venture capital format to encourage investment: they stipulate that, with any investment, one-third goes towards buying stock and two-thirds go towards a low-interest, long-term loan to the company. "There are few investment banks in China, but the principle is the same," says Mao. "We just borrow from individual investors instead of investment banks." For the investors, this lowers the financial risk, should the company

fail. "Normally they would lose their entire investment, and we would only lose our technology. Now, they would only lose one-third of their investment, and we would still owe them the rest," says Mao. "It shows that we are confident in our products."

Apart from the struggle to generate a favourable climate for biotech investment, the other main obstacle to China's development in biology is the eccentricity of funding decisions, which are often based on political leverage rather than peer review. The result, critics claim, is a proliferation of uncoordinated projects. "Everybody is just trying to jump on the genomics bandwagon, but very few know where they are going," says Ming-Wei Wang, formerly chief executive officer of a Shanghai-based biotech company. "In China, everything is decided in backrooms, and then just announced," says one researcher with the Chinese Academy of Sciences. Genomics research in China is an uncoordinated "bunch of loose ends", complains another.

But others are optimistic that China will begin to apply more rigorous, open and accountable methods of decision-making. If so, its future as a leading international player in biology should be secure. Li Jin, formerly of Fudan University's Institute of Genetics, and now based at the University of Texas at Houston, is one of those predicting a bright future. "With increasing funding and by taking advantage of its unique resources, China will soon become a serious contender on the world stage," he says. ■

David Cyranoski is *Nature's* Asian-Pacific correspondent.

1. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
2. Xia, J.-H. et al. *Nature Genet.* **20**, 370–373 (1998).
3. Yang, X. et al. *Am. J. Hum. Genet.* **66**, 892–903 (2000).
4. Zhang, X. et al. *Nature Genet.* **27**, 151–152 (2001).
5. Xiao, S. et al. *Nature Genet.* **27**, 201–204 (2001).

Breaking the mould

In choosing an outspoken former government science adviser as its president, the Royal Society has departed from tradition. Peter Aldhous spoke to Robert May about his plans for Britain's national scientific academy.

March 1999 was a tough time to be chief scientific adviser to the British government. An intensive campaign by environmentalists to convince the public that genetically modified (GM) crops were unsafe was in full flow. The green lobby provided the ammunition as competing tabloid newspapers rivalled one another in their anti-GM fervour.

Robert May did not take a low profile. He described the papers' coverage as "crap", and was roasted by one editor for treating her readers with contempt. "It was, for me, a very steep learning curve, and I made some very instructive errors," May says.

The incident did not surprise May's friends. He is never afraid to argue his point, and often does so in colourful language. So when it emerged last year that May would be the next president of the Royal Society, a few eyebrows were raised. Britain's scientific academy is seen as a conservative pillar of the Establishment. May, with his blunt manner and high media profile, did not fit the mould.

Going public

In fact, May's appointment is the most obvious move in an ongoing drive by the Royal Society to engage in public debate. In Britain, trust in scientists has been severely damaged by the revelation that bovine spongiform encephalopathy can be transmitted to people, and further dented by the skilfully orchestrated backlash against GM food.

In this charged atmosphere, May's experience of working inside government, and the lessons learned from his time in the media spotlight, could be valuable. And no one doubts his scientific credentials. Trained in his native Australia in theoretical physics, he has brought his formidable mathematical skills to bear on ecology, conservation biology and the dynamics of the struggle between viruses and the immune system — and he still maintains a position at the University of Oxford.

Since taking over at the Royal Society on 1 December, May has been uncharacteristically quiet. But last week, after chairing his first meeting of its governing council, he was happy to talk about his plans. Pitching the Royal Society into the centre of debate over scientific controversies is top of his agenda. He hopes to build on the society's recent

success in briefing Members of Parliament about stem-cell research in the run up to their vote on allowing embryos to be used to investigate therapies for disease (see *Nature* 409, 5; 2001).

Meeting of minds

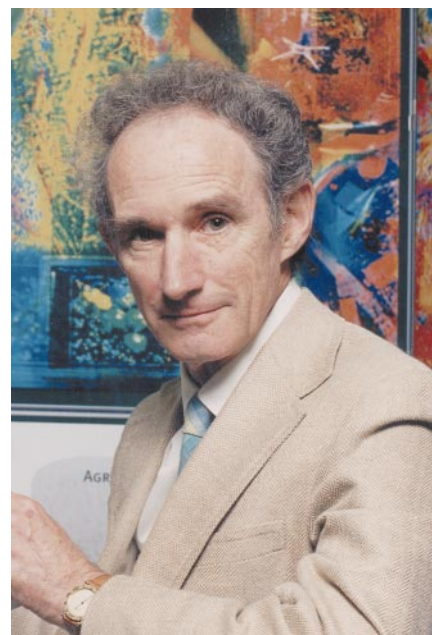
Another tactic will be to repeat the format of an unusual meeting held last year. The Royal Society holds regular two-day discussion meetings, which usually attract little attention outside the fields concerned. But last September, the society's London headquarters bristled with television cameras as experts on HIV debated a controversial theory, proposed by the British journalist Ed Hooper, that AIDS started in the 1950s through the distribution of a polio vaccine cultured from contaminated chimp tissues (see *Nature* 407, 117; 2000).

That meeting happened because of the persistence of its promoter, the late Bill Hamilton of the University of Oxford. Hamilton, a leading evolutionary biologist who was sympathetic to Hooper's ideas, contracted malaria while in Africa seeking evidence to test the theory, and died before the conference took place. Some scientists were concerned about the meeting degenerating into a media circus, but it is now recognized to have done much to bring the underlying scientific issues, and the paucity of evidence for Hooper's theory, to public attention.

"Some felt there was an air of vulgarity," May says. "But I thought it was an example of exactly what the Royal Society ought to be doing." In future, he wants a couple of slots held open each year to schedule meetings at short notice on matters of public concern.

Fellow fever

If the Royal Society is to engage with the public more effectively, May argues that it must also be more inclusive. Just 42 of its more than 1,000 fellows are women, and those working in non-traditional disciplines are also under-represented. May cites Tim Berners-Lee, who invented the World-Wide Web while at CERN, the European Laboratory for Particle Physics near Geneva, as an example. Last year, Berners-Lee was awarded one of the society's prestigious Royal Medals. "But he's not a fellow, because no one had



If you want to be more effective in engaging issues of public concern, then you really need to understand the rules of engagement.

thought to nominate him," May says.

To help to ensure that such pioneers do not get overlooked, the society's council has now endorsed a scheme that will no longer require nominees to be proposed by six existing fellows. In future, a proposer and a seconder will do. The scheme was initiated before May's arrival. And he stresses that most of his plans merely represent an extension of the policies of his predecessor, Nobel laureate Aaron Klug of the Laboratory of Molecular Biology in Cambridge.

But in appointing a high-profile former government adviser, the Royal Society is intensifying its moves into the public and political arena — and is taking a calculated risk. "If you want to be more effective in engaging issues of public concern, then you really need to understand the rules of engagement," says May. "But you must always worry that there's a conflict of interest, or that you're bringing baggage from a previous job."

May also accepts that some traditionalists feel uneasy about the appointment of someone with his reputation for straight talking. "There were those who felt that I'm insufficiently dignified," he says. But, true to form, he is not too worried about causing offence. "With respect to the interaction with some people, it will make things easier," he smiles. ■

Peter Aldhous is *Nature's* chief News and Features editor.

♦ <http://www.royalsoc.ac.uk>

All sectors of society must work together to save biodiversity

Sir—Conservation International (CI) is an organization that “protects the Earth’s biologically richest areas and helps the people who live there improve their quality of life”¹. Globally, we focus on those areas richest in irreplaceable biodiversity: hotspots² and tropical wilderness areas³.

A recent survey⁴ of 31 expatriate and 74 Indonesian environmental professionals found that most of them were more concerned with sustainable development and land-use planning than with species and wilderness protection.

The conclusion was that “conserving species and tropical wilderness areas will require that policy-makers in targeted regions give CI’s ecological goals higher priority than is currently the case in Indonesia”⁴.

We fully agree with the paper’s findings and conclusions. However, we should point out that it is important not to confuse geographic global conservation priorities with strategies for conservation implementation within these priorities. The broad-scale focus of CI, and of the Critical Ecosystem Partnership Fund with the World Bank and Global Environment Facility⁵, on hotspots and tropical wildernesses, concerns the former. In contrast, the Indonesia survey⁴ concerns the latter.

Our primary focus at this fine scale is on strengthening direct protection of biodiversity. The greatest current concern is that Indonesian internal conflicts and economic crisis might deteriorate so far that biodiversity conservation becomes a low priority. However, the current political crisis and forest management decentralization process has created an atmosphere in which significant shifts in conservation strategy are being seriously considered. Innovative ideas are welcome, such as the payment of local opportunity costs⁶ via the ‘conservation concessions’ concept⁷. Nevertheless, we would argue that there is no single strategy for conservation implementation — rather, the appropriate actions must be determined by local knowledge of the situation.

The current CI–Indonesia programme, for example, is involved in activities that support sustainable development programmes and help integrated land-use and biodiversity planning in Irian Jaya, the easternmost province in Indonesia. In central Sulawesi, we are creating economic alternatives to deforestation and other

habitat loss by developing ecologically sustainable enterprises such as ecotourism in the Togeian proposed marine park, and we are supporting Non-Timber Forest Product development in Sumatra.

To sum up, now is the time for conservation organizations such as CI to secure environmental stewardship and biodiversity conservation goals in partnership with the Indonesian government, NGOs, the university community, business and civil society. The fate of one of the planet’s richest ‘megadiversity’⁸ countries is at stake.

Jatna Supriatna

*Conservation International—Indonesia,
Jl. Taman Margasatwa 61, Jakarta 12540,
Indonesia*

1. <http://conservation.org/about.htm> (accessed 4 January 2001).
2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. *Nature* **403**, 853–858 (2000).
3. Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, G. A. B. & Olivieri, S. *Conservation Biology* **12**, 516–520 (1998).
4. Jepson, P. *Nature* **409**, 12 (2001).
5. <http://www.cepf.net/> (accessed 4 January 2001).
6. James, A. N., Gaston, K. J. & Balmford, A. *Nature* **401**, 323–324 (1999).
7. Gullison, R. E., Rice, R. E. & Blundell, A. G. *Nature* **404**, 923–924 (2000).
8. Mittermeier, R. A., Robles Gil, P. & Mittermeier, C. G. *Megadiversity* (CEMEX, Mexico City, 1998).

High rate of inbreeding in Spanish universities

Sir—Over the past couple of years *Nature* has hosted an intense debate on the state of Spanish science (see, for example, refs 1–3). Two issues are repeatedly raised: the lack of sufficient funding and the existence of social networks that, regardless of the candidates’ scientific merit, systematically award positions to one of their members. While the former is *a priori* easily quantified^{1–3}, allegations of “blatant endogamic practices”² have thus far remained untested.

An indication of inbreeding can be obtained by determining whether the address of a scientist’s first publication coincides with their current address as a faculty member. Using this measure, we compared possible inbreeding in Spanish universities with the situation in three other countries: the United States, the United Kingdom and France.

We collected data from the Web of Science (ISI, WoS version 4.3, <http://wos.mimas.ac.uk>) last October on 160 randomly sampled researchers — 40 from each country — holding permanent faculty positions in science departments. In Spain they held the position of ‘profesor titular’, in the UK ‘lecturer’, in the United States ‘assistant professor’ and in France ‘maitre de conference’.

In order to determine that the first publication to appear in WoS (which spans

1981–2001) was indeed the first paper published by the researcher, we only included individuals whose first recorded publication was in 1984 or later. To avoid biases caused by recently created universities, where all the faculty are necessarily external, we only considered faculty working in universities that had existed for more than 50 years.

The percentage of external candidates that obtained a permanent faculty position in each country is revealing: 93% of candidates to posts in the United States were external, as were 83% in the UK and 50% in France. In Spain, by contrast, only 5% of lectureships were given to individuals who had published their first paper while working in another institution. Differences are very significant (Kruskal–Wallis, $p < 0.0001$), the percentage registered in Spain being at least ten times lower than in the other countries.

These observations are consistent with allegations that lectureships at Spanish universities are almost exclusively awarded to individuals who started their scientific careers in the same institution. Alternative explanations, such as a reluctance among Spanish researchers to move from a geographic region, are unlikely as, in such a scenario, some degree of exchange of scientists between institutions within the same region would have been expected. Spain’s Ministry of Education currently lists more than 60 universities (<http://www.mec.es>) and the Spanish Research Council (<http://www.csic.es>) lists more than 90 science-related research institutes spread across the country.

Our observations thus suggest that Spanish universities are almost completely impermeable to external candidates, effectively preventing the movement of researchers and thus the exchange of ideas and expertise which is one of the keystones of scientific progress.

Attempts to revitalize the state of Spanish science are being made by both Spanish scientists and institutions. These include such welcome initiatives as the Manifesto for a Social Pact for Science and Technology⁴, and innovative plans by the Catalan government to create research contracts with periodical evaluations of scientific performance. But these efforts will be wasted if researchers are employed on grounds other than their scientific merit.

Arcadio Navarro, Ana Rivero

*Institute of Cell, Animal and Population Biology,
University of Edinburgh, Kings Buildings, West
Mains Road, Edinburgh EH9 3JT, UK*

1. *Nature* **407**, 659 (2000).
2. Ferrer, P. *Nature* **407**, 941 (2000).
3. Marimon, R. *Nature* **408**, 18 (2000).
4. *Nature* **408**, 397 (2000).

Highlighting the grey matter

An attempt to answer the perennial questions of neuroscience.

What Makes You Tick?

The Brain in Plain English

by Thomas B. Czerner

Wiley: 2001. 228 pp. \$24.95, £16.50

Steven P. R. Rose

Neurobiology may be the fastest growing of contemporary sciences, but interpreting for a lay audience the complexity of what we are discovering about the brain remains a major challenge — not least because the questions that rightly concern those outside the field, classically those of the relationship of mind and brain, and of neural processes and conscious experience, are just those that we neurobiologists have the greatest trouble with ourselves.

A few years ago, I agreed to write a book for children about the brain, jointly with a 12-year-old friend. Asked what he wanted to know about the brain, my co-author polled his classmates and came back with a list of 13 questions, the last two of which read: 'how does your brain make you happy or sad?' and 'are you the same as your brain, or different?'. We were allocated 32 pages, plus some fine illustrations, in which to try to answer these abiding philosophical questions for kids. Thomas Czerner has rather more space, but minimal and somewhat impoverished visuals, in which to do the same for adults. His book has a slightly breathless style, which is sometimes hyperbolic with enthusiasm.

As all neuroscientists know, the questions we ask about the brain range from the molecular to the systemic. And in the gulf between our lab practice and coffee-room talk, we oscillate between a simplistic reductionism and a naive dualism. We divorce our personal, subjective experience from the objectivity of analysing the cells and molecules we manipulate. Czerner's book exhibits this tension nicely.

Early on, he distances himself from Francis Crick's claim that "you're nothing but a pack of neurons" by insisting that "in fact, you are not merely a pack of cells of any kind". By this he seems to claim a sort of dualism: a person is an entity containing a hundred billion neurons and trillions of synaptic impulses; and this pack of neurons "produces a peculiar music that you can hear, feel, taste and smell". Elsewhere, he argues that subjectivity is an emergent property.

But Czerner can't sustain this dualism for long, and we are soon into the less intellectually problematic nitty-gritty of cell membranes, G proteins and NMDA receptors. Indeed, by the end of the book, he seems to relapse fully into a genetic reductionism —



Mind set: a hundred billion neurons produce a peculiar music that you can hear, feel, taste and smell.

violence is a result of low serotonin levels partially caused by a deficiency in monoamine oxidase A, tranquillity is a product of adequate tryptophan metabolism. Gung-ho triumphs as it appears that genetic engineering will soon cure Huntington's disease and enhance learning and memory. No hint of criticism or doubt, either scientific or ethical, is allowed to cloud this glorious, just-over-the-horizon prospect. It would have been interesting had Czerner reflected on the seeming incompatibility of this dualist approach to brain, mind and human agency. But despite an engaging preface describing the trajectory of his personal life, he evades the task.

So how should one present the brain? Classically, one begins with the gross anatomy and crude functional divisions, then drops down into synaptic neurophysiology and some molecular mechanisms, and finally emerges into the world of behaviour and altered brain states. As an ophthalmologist, Czerner's approach is somewhat different. There's a scrap of evolution, with a rather old-fashioned thesis about how the 'human' neocortex is superimposed on an ancient, reptilian limbic system, but not much about brain development. Molecular mechanisms come and go, although, I suspect, not described in enough detail to help a non-biochemically savvy reader. Interspersed with the straight state-of-the-art story, there are snippets of history and

personal anecdote. Heroes include Stephen Kuffler, George Wald and Julius Axelrod. As can often be the case in books emanating from the United States, the NIH (Not Invented Here) syndrome is alive and well; whatever the past achievements of Europeans, they don't feature among the present-day movers and shakers.

In terms of how to choose what to cover and what to leave out, not surprisingly Czerner focuses on the visual system. This makes good sense not only because of his background, but also because visual processes are the best understood of sensory systems. But I found a strange lack of system or logic in the constant switches of level within the text.

The crucial test, of course, is not what a neuroscientist thinks, but how well and accurately the book communicates with its intended audience. An interesting comparison is with the recent tours of the brain conducted by the British neurobiologist Susan Greenfield. Czerner shares a good deal of Greenfield's infectious enthusiasm and optimism, but comes across as more of an outsider, less prepared to make critical judgements (as on the relationships between genes and behaviour, for instance). He writes more simplistically and speculates more freely. Moreover, the science is spiced up with anecdotal accounts of researchers and, as can often be the case in neuroscience and genetics, patients. Not surprisingly, then, we

book reviews

find the unfortunate Phineas Gage — who survived having a three-foot, seven-inch iron rod blasted through his brain — playing an early and prominent part in Czerter's story. I would have preferred the book's treatment to have been a little more restrained and to have demanded more of its readers. ■

Steven P. R. Rose is in the Brain and Behaviour Research Group, Open University, Milton Keynes MK7 6AA, UK.

The real benefits of copycats

The Imitation Factor: Evolution Beyond the Gene

by Lee Alan Dugatkin

Free Press: 2001. 256 pp. \$25

Stephen Pruett-Jones

The image of a chimpanzee eating a banana seems so iconic of what it means to be a chimpanzee that I have always assumed they instinctively knew how to eat a banana. In fact, I was wrong. Chimpanzees must learn that bananas are appropriate food items and, through imitating others, how to open them. Not only that, they even have to learn how to recognize most food items, what predators to avoid and how to interact with other chimpanzees.

In this respect, chimpanzees are not unique. As Lee Dugatkin reminds us in *The Imitation Factor*, learning through imitation, teaching and the cultural transmission of traits is not limited to humans. It is common in many species, playing a critical role in their lives. To convince a lay reader — the book's intended audience — of the importance of cultural transmission in animal societies is a laudable goal, and I would have recommended this book highly had Dugatkin stopped there. Unfortunately, he didn't.

The Big Idea of the book — and perhaps one needs big ideas to sell popular science books — is that not only is imitation important in animal societies, it is also a major influence in biological evolution (an “unrecognized force of nature”) and “how evolution proceeds beyond the gene”. So strong is the influence of cultural transmission on evolution, Dugatkin argues, that this influence partially rehabilitates Lamarck's theory that characteristics acquired by an organism during its lifetime can be passed on to its descendants — “although Lamarck was incorrect about acquired characteristics and genes, he was not wrong about acquired characteristics and evolution”.

By extending this argument to culturally transmitted traits, Dugatkin is easily led to exaggeration. “Given what we know about



Who's copying whom? It remains to be seen whether imitation significantly affects evolution.

culture, Crichton's example, under the right conditions, could happen”. This quote refers to the 1995 movie *Congo*, based on Michael Crichton's novel, in which gorillas are taught to use weapons to protect a diamond mine and then teach their own offspring to use them long after the miners are gone.

The hypothesis that imitation can influence biological evolution goes something like this. Imagine a species in which some females do not choose their mates independently, but rather ‘copy’ the mate choice decisions of other females. Such mate copying influences the distribution of matings among males, and thus changes the distribution of genes passed on to the next generation. Hence, imitation affects evolution. Theoreticians have documented the plausibility of this scenario, and there are several well-documented cases from nature. In the case of Darwin's finches, for example, songs are culturally, not genetically, inherited. Because females use the song of males when deciding on a mate, a culturally transmitted

trait influences not just species recognition, but also species formation.

But problems arise when Dugatkin extrapolates such evidence to suggest that imitation acts independently of genes and generates broad evolutionary change. First, females who copy the mate choice of others (or individuals who imitate the food choices of others) are ‘copying’ the actions of individuals whose own choice patterns are, at least in part, genetically determined. Thus, in general, imitation will enhance the natural or sexual selection process. Furthermore, although an imitated behaviour may not be under direct genetic control, the flexibility of the behavioural repertoire that leads to cultural transmission in the first place is influenced by genetics. In the Darwin's finches example above, although the specific song that individual nestlings learn is culturally determined, the ability to learn songs and the sensitive period during which nestlings can learn songs are genetically determined. Thus, imitated behaviour

and genes are always connected at some level.

Second, with respect to female copying, in very few species have females actually been shown to copy one another in their choice of mates, and one of the best-studied cases — Dugatkin's own work on Trinidadian guppies — is being questioned because two independent attempts to replicate the basic study have yielded negative results.

Third, in non-human species, natural selection acts as a very tight filter through which all observable traits must pass. Animals can be trained to do many things in the lab — for example, learn inefficient or maladaptive traits and culturally transmit them to others. But in the wild, natural selection quickly eliminates most maladaptive traits, however they are passed on.

Finally, it is important to remember that the vast majority of animal species are either not sufficiently complex for the evolution of culture to be relevant, or are solitary in nature, so that individuals seldom experience the conditions necessary for imitative behaviour to occur — as is the case, in fact, for most mammals.

Imitation clearly plays an important role in cultural evolution and, in limited cases, biological evolution. Nevertheless, it usually acts in concert with natural selection, its scope is limited, and the fact that cultural evolution is important in animal societies is hardly new. This book will provide a lay reader with some interesting anecdotes about animals, but I doubt that it will give them a better understanding of evolution. ■

Stephen Pruett-Jones is in the Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA.

Evolutionary celebrities

The Cichlid Fishes: Nature's Grand Experiment in Evolution

by George W. Barlow

Perseus: 2000. 335 pp. \$28

Axel Meyer

David Star Jordan, the eminent ichthyologist, anti-darwinist and first president of Stanford University, is said to have justified his refusal to memorize students' names with the excuse that, for every student's name he was supposed to remember, he would forget the name of a fish species. That was in the days before student evaluations of professors.

Unfortunately for students, there are many fish names to remember. The 25,000 or so living species of teleost fishes are systematically arranged into more than 400 families. And, to paraphrase Thomas Huxley, God must have had a particular fondness for the

Cichlidae family, as it contains perhaps up to 15% of all species of fishes.

The most ancestral cichlids are found in Madagascar and India, and hundreds of cichlid species live in Central and South America. But the centre of biodiversity is East Africa. In particular, Lakes Victoria, Malawi and Tanganyika each harbour species flocks of hundreds, possibly more than 1,000, endemic species. None of the many other families of fish living in these lakes comes even close to rivalling the cichlid's dominance in terms of species diversity.

Only since the advent of molecular phylogenetic techniques, which can trace the evolutionary development of species at the molecular level, has it become known that all of these diversifications can be traced back to a single ancestral lineage — or at least a very small number — that colonized each of the lakes. One implication of this is that all the incredible ecological, morphological and behavioural specializations that are found in these three groups of cichlids arose independently of one another. This would make these species flocks one of the most spectacular examples of convergence in all of evolutionary biology.

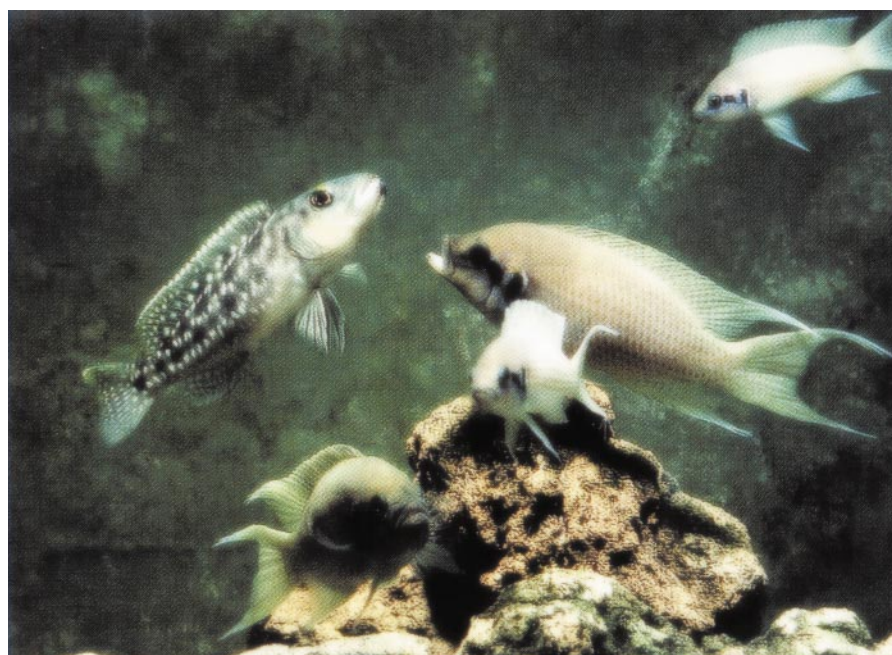
Cichlid taxonomists had assumed that if two species in different lakes showed very similar morphologies or behaviours, then they were closely related. Who could blame them? For it would have seemed preposterous to assume that evolution was so lazy or unimaginative as to reinvent such unbelievable diversity three times over. Moreover, molecular data also show that the speciation and diversification of cichlids took place in record time — in Lake Victoria around 500 species arose in probably less than 14,000 years. Cichlids are evolutionary

celebrities indeed. In case you don't remember, Wanda — she was the piscine star in the movie *A Fish Called Wanda* — was a cichlid, a South American species called the angel fish.

Ecologically, cichlids seem to employ every possible trick. But if hundreds of often extremely closely related species coexist in one lake, they have to keep out of one another's way competitively. Some are specialized to feed only on the scales of other cichlids, some use the specialized second set of jaws — probably part of the explanation for the cichlids' evolutionary success — to crack open snail shells. Other species suck the young out of the mouths of brooding cichlid mothers, and still others tend the gardens of algae on which they feed and which they vigorously defend against other cichlids. Many of these astonishingly specialized adaptations are found repeated in the species flocks in all three lakes.

Cichlids are smart, and in terms of mating systems and parental care systems, they are more diverse than any other family of fish. They have been the white rats of fish behavioural research for more than 50 years. George Barlow, the author of *The Cichlid Fishes*, has dedicated his life's work to cichlids. He single-handedly made the University of California at Berkeley the premier centre of research in cichlid behaviour during the heydays of animal behaviour research in the 1970s and 1980s.

In this book, which is written in a colloquial and non-technical style, Barlow summarizes decades of his experience with cichlids. Fishy tales are often interwoven with human anecdotes. The focus is on many of the fundamental (and often still unsolved) questions in behaviour and sociobiology. Communication, aggression,



Fishy business: the Cichlidae family at home in Lake Tanganyika.

mating systems, parental care systems and parent-offspring conflict are all covered: the list seems all too human. Barlow has a behavioural view of the world, and draws on his impressive knowledge of the behaviour and natural history of many groups of fishes to make more general, important points about evolution and the ecological circumstances that shape the development of parental care and mating systems.

But Barlow has aimed this book at a wider, not strictly hard-core scientific, audience. There is a huge community of what he calls "cichlidiots" who keep and breed the typically beautifully coloured and 'brainy' cichlid fishes. They will enjoy the lucid writing and the plethora of general biology facts and aquarium behavioural observations. The book is also entertaining and fun to read because of (or in spite of, depending on your taste) headings in chapters such as "Beauty is only fin deep", or statements such as "gonads are in command".

Cichlid Fishes has a different focus from the grand ol' book of cichlid research, Geoffrey Fryer and T. D. Iles' 1972 text, *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution* (Oliver & Boyd). The latter book focuses more on ecological and evolutionary questions, and, owing to very active research by many laboratories around the world, is quite outdated in places. An updated synthesis focusing specifically on the ecology and evolutionary biology of cichlids is still wanting. ■

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

Science in culture

Architect reaches for the clouds

How fractals may figure in our appreciation of a proposed new building.

Richard Taylor

We're all familiar with the Manhattan skyline of New York, with its many skyscrapers reaching into the clouds. Imagine the effect if one of these skyscrapers were shaped like the clouds surrounding it. This is the plan of the Guggenheim Museum, which recently unveiled a design by the architect Frank Gehry for an \$800 million building to house its modern art collection. The Guggenheim Museum is no stranger to controversial architecture. Its current collection is housed in the landmark white spiral structure designed by Frank Lloyd Wright. But with its swirling layers of curved surfaces spanning three piers, the proposed 45-storey, 'cloud-like' structure is expected to radically reshape New York's waterfront.

Traditional architecture is based on euclidean shapes, such as circles, squares and triangles. But clouds belong to fractal geometry, consisting of patterns that recur at increasingly fine magnifications. How, then, will people react to architecture designed to mimic one of nature's fractal patterns? The answer may lie, not in architectural concepts, but in recent perception studies of fractal patterns.

The study of human aesthetic judgement of fractal patterns constitutes a relatively new research field of perception psychology. Only recently have researchers started to quantify people's visual preferences for (or against) fractal content. The visual appearance of a fractal object is influenced by a parameter called the fractal dimension, D . This quantifies the fractal scaling relationship between structure observed at

different magnifications. Its value lies between 1 and 2 and moves closer to 2 as the complexity and richness of the repeating structure increase.

In 1995, Cliff Pickover at the IBM Thomas J. Watson Research Center in New York used a computer to generate fractal patterns with different values of D (ref. 1). He found that people expressed a preference for fractal patterns with a value of 1.8. A subsequent survey by Deborah Aks and Julien Sprott at the University of Wisconsin also used a computer, but with a different mathematical method for generating the fractals². This survey reported much lower preferred values of 1.3. The discrepancy between the two surveys seemed to suggest that there is no universally preferred D value, but that the aesthetic qualities of fractals instead depend specifically on how the fractals are generated.

To determine whether there are any 'universal' aesthetic qualities of fractals, I collaborated with psychologists Branka Spehar at the University of New South Wales, Colin Clifford at Macquarie University in Sydney and Ben Newell at University College London. We performed perception studies incorporating the three fundamental categories of fractals — 'natural' fractals (scenery such as trees, mountains and clouds), 'mathematical' fractals (computer simulations) and 'human' fractals (cropped sections of Jackson Pollock's dripped paintings, which I have shown to be fractal³). Participants in the perception study consistently expressed a preference for fractals with D values in the range 1.3 to 1.5, irrespective of the pattern's origin. Significantly, many of the fractal patterns surrounding us in nature have D values in this range. Clouds have a value of 1.3.

Although Gehry's proposal for the Guggenheim Museum is designed to mimic the general form of clouds, it is clear that the completed building will not be strictly fractal. To build a structure described by a D value of 1.3 would require many layers of repeating patterns. Although this is no great challenge for nature, such complexity is beyond current building techniques. In fact, both Gehry and New York's mayor, Rudolph Giuliani, readily admit that no shovel will be turned for at least five years and that the plans will have to evolve during that time. It will be fascinating to see if people's fundamental appreciation of fractal clouds will inspire New Yorkers to embrace this revolutionary building design. ■

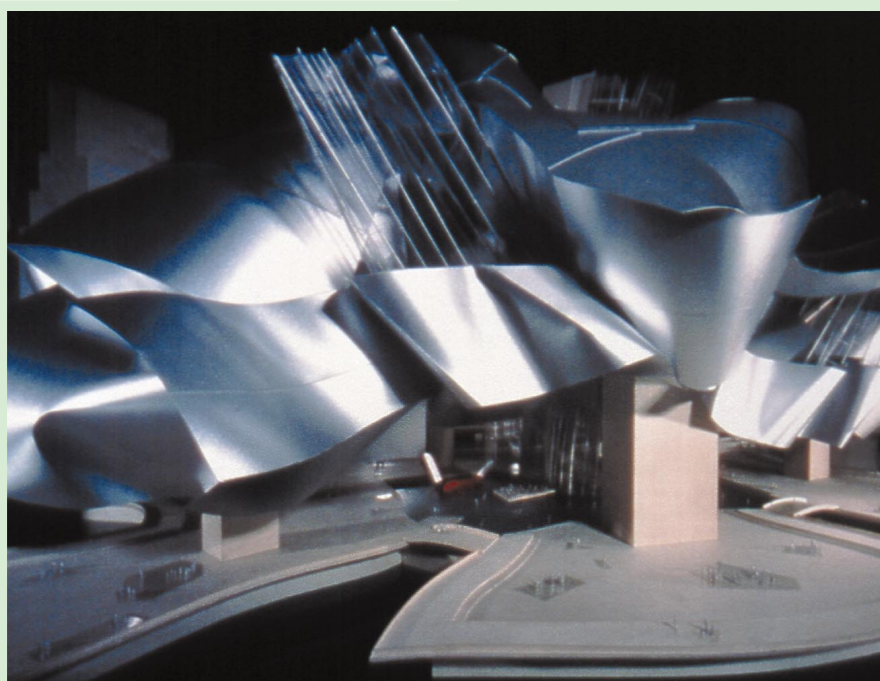
Richard Taylor is in the Department of Physics, University of Oregon, Eugene, Oregon 97403-1274, USA.

1. Pickover, C. *Keys to Infinity* 206 (Wiley, New York, 1995).

2. Aks, D. & Sprott, J. *Empir. Stud. Arts* 14, 1 (1996).

3. Taylor, R. P., Micolich, A. P. & Jonas, D. *Nature* 399, 422 (1999).

Fractal future: a model of the proposed new Guggenheim Museum in New York.



DAVID HEALD

Character-building

How do writing systems such as China's deal with the twenty-first century?

Alan L. Mackay

The alphabet's place as the basis of European culture is so fundamental an assumption that in considering the role of the word — "In the beginning was the Word" — we may forget that other people have different views and that Faust countered with: "In the beginning was the deed." After all, the alphabet provides us with powerful analogies for science. Two millennia before the genetic code was broken, Epicurus observed: "The atoms come together in different order and position, like the letters which, though they are few, yet, by being placed together in different ways, produce innumerable words."

Yet the Chinese writing system — of characters that represent a word's meaning, not its sound — has survived for several millennia, to the extent that some glyphs of 3,000 years ago are still understandable today. While other systems deriving from pictographs, such as those of the ancient Middle East, succumbed to competition from alphabetic and phonetic systems, the single set of characters used throughout the empire to represent all China's many different dialects has served as a powerful unifying administrative tool.

The typewriter, the printing-press and the electric telegraph were absorbed in China with the help of some not-very-satisfactory technological fixes, and the need to use computers might have put an end to the traditional writing system. Instead, the computer has come to the rescue and Chinese characters have a new life. It is a matter of amazement to visitors from the alphabetic world that a modern civilization can be conducted with this system, which even copes with voice-recognition software. In addition to the traditional shopping streets in Beijing selling jade and silks, there is now a long street of computer shops. Even pedal rickshaws have been adapted in the district to enable purchasers to bring home the large boxes containing their new computers.

How does the Chinese writing system cope with the new concepts regularly thrown up by science, such as plutonium and the Internet? Despite many millennia without the alphabet, China was, in the early 1950s, on the point of giving up the traditional characters in favour of a phonetic alphabet to meet twentieth-century needs. In fact, although a phonetic alphabet, *Pinyin*, is now universally taught as a supplement, characters continue to be primary. Older people make mistakes with *Pinyin*, but children are

absolutely accurate. New concepts get new words and sometimes new characters.

They adapt in a way that C. P. Snow, who spoke tellingly of the growing gulf between the literary and scientific worlds, would have admired. Snow used to go around cocktail parties asking people if they could explain the Second Law of Thermodynamics. "A good many times I have been present at gatherings of people who, by the standards of traditional culture, are thought highly educated and who have with considerable gusto been expressing their incredulity at the illiteracy of scientists," he said in his 1959 Rede Lecture "The Two Cultures and the Scientific Revolution". "Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold; it was also negative."

It happens that there is an exactly corresponding test for the Chinese cultural area which also divides the literary and scientific populations. Ask your Chinese friends if they know the character shown above.

A Chinese character usually has two parts. On the left or above is the 'radical', which gives some hint of the meaning. Characters are to be found in the dictionary by looking up the radical, of which there were traditionally 214 (although the inventory has been modified), and then counting the brush strokes necessary to write the remainder. The part beside the radical is usually called the 'phonetic', and gives some guide to the pronunciation. Old components are used in new combinations.

Thus the character in the figure has the fire radical on the left and on the right it has, as phonetic, the character for 'merchant', which is pronounced *shang*. The character as a whole means entropy, and is also pronounced *shang*. I suppose that it was created by a committee and it now occurs in the international standard matrix of characters usually called GB-2312, which is known to all Chinese word processors (though not to all people). In computer systems each character is coded by two ASCII bytes or 16 bits;

The computer has come to the rescue and Chinese characters have a new electronic life.



A source of confusion: entropy.

there are some 6,000 Chinese characters plus the main alphabets in the standard font.

People remember characters by their real or imagined etymologies, so that the association of fire or energy with buying and selling is a very appropriate mnemonic for the way to write 'entropy'. In a more entropy-conscious world, if energy were to be priced according to the temperature at which it is to be delivered, people would not use electricity for space-heating.

In spite of appearing to be a stream of single characters, the Chinese language consists mainly of polysyllabic words made up of several characters. With a modern word processor, if you type in a single syllable, the program will offer you perhaps a dozen characters all pronounced in the same way. You could choose one of them, but if you continue to type successive phonetic syllables of the word or words in a phrase, there is usually a unique output of characters. Entering the phonetic *shang* brings up eleven characters, with 'entropy' as the least likely choice.

The enunciation of the phonetic *shang* by itself is thus not enough to bring up the meaning 'entropy', so a teacher has to write the character on the board or, in conversation, launch some explanation of what is being talked about. Nevertheless, at a Chinese version of C. P. Snow's cocktail parties, people can talk about whatever may be necessary, exploring each other's knowledge or ignorance, and later repair any deficiencies using the Chinese Internet. Thus the Chinese language and its writing system continually prove to be adaptable to changing requirements and have responded creatively to the challenge of electronic communication. ■

Alan L. Mackay is at the School of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

Perceptions of knowledge

Richard L. Gregory

Living organisms are leaky test tubes of reactions. Almost isolated from the rest of the Universe, they receive limited information for adaptive behaviour, prediction and planning. The 'proximal' senses of taste and smell directly monitor chemical interchanges, and touch detects contact and damage. But the 'distance' senses of vision and hearing are very different. They provide only indirect knowledge of what matters — requiring interpretations from knowledge and assumptions, so you can read meaning into the object world.

One cannot eat, or be eaten, by the optical images in eyes. Images are useless shadows of objects until their significance is read. This process is so complex that even the most advanced artificial-intelligence computer programs are far from replicating it. Reading objects from images requires knowledge of the world of solid, interacting things. When this knowledge is not appropriate, the eyes' images may be misread, giving 'cognitive' illusions. An example is distortion in pictures that show depth by perspective. Our response to

perspective shrinking means our brains expand features represented as distant to about the correct size, even though they lie flat on the picture plane. When the visual scaling is inappropriate, distortions are generated.

The knowledge we use when we see has come from millions of years of interacting with objects. At first, lessons of survival, such as reflex behaviour, were stored in and inherited from the genetic code. Then knowledge was gained individually, largely by pseudopodia-to-hands-on exploration. As knowledge is so important, tricks of camouflage and deceit became potent biological weapons. Surely science is itself a remarkable extension of millions of years of discovery, making new sense of sensory signals. With added data from instruments, science develops general concepts that are extremely different from ancient object-knowledge. So we live in two worlds — perceptions of experience alongside conceptions of understanding — both based on knowledge and assumptions that may be wrong.

It is not easy to define 'illusion'. Illusions are departures from truths of the object world: but how can we know what these are? There are very different claims of truth by science, art, religions and the many flavours of metaphysics. Which should we accept as reference reality for recognizing illusions? Science has changed its concepts of reality many times, and science's realities grow ever further from how things appear. So we may be tempted to think of all perception as illusion. But this is no more helpful than claiming that reality is a dream — when the word 'dream' loses any meaning, and so does 'illusion'.

The familiar visual illusions in children's books and psychology texts are departures from simple measurements of lengths, curvatures and so on, of the common-sense world of appearances. Let's call this common-sense reality 'kitchen physics'. Illusions are measured by matching appearances against this common-sense 'kitchen' world, using rulers, scales, thermometers, clocks and so on, as found in kitchens. There is no reference to concepts of

physics for defining these illusions. Yet no doubt quantum physics is essential for understanding cooking.

The main lesson of illusions is that perceptions are not tied to object reality. Perceptions are guesses — predictive hypotheses — of what may be out there. They are our most intimate reality; yet as for any hypotheses, they may be wrong — especially when based on false assumptions, or generated by misleading procedures, when even visual paradoxes can be created.

Illusion

"We live in two worlds — perceptions of experience alongside conceptions of understanding — both based on knowledge and assumptions that may be wrong."

There are not only qualitative illusions, such as distortions, but also qualitative phenomena: especially objects spontaneously transformed into other objects, such as ducks turning into rabbits. Alternative visual hypotheses are entertained, in turn, when the brain seeks a better answer. The alternative perceptions are drawn from the inner scene of the brain; so these dynamic ambiguities reveal something of mind. They are indeed insights.

The main illusions can be classified as ambiguities, distortions, paradoxes and fictions. Perhaps these correspond to errors of language, for language might have developed from pre-human perceptual classifications of objects and actions. Could this be the 'deep structure' of language — perceptions by past species, of lost worlds, lying deep in our vision and speech today? No wonder things look confusing if our mental maps are millions of years out of date.

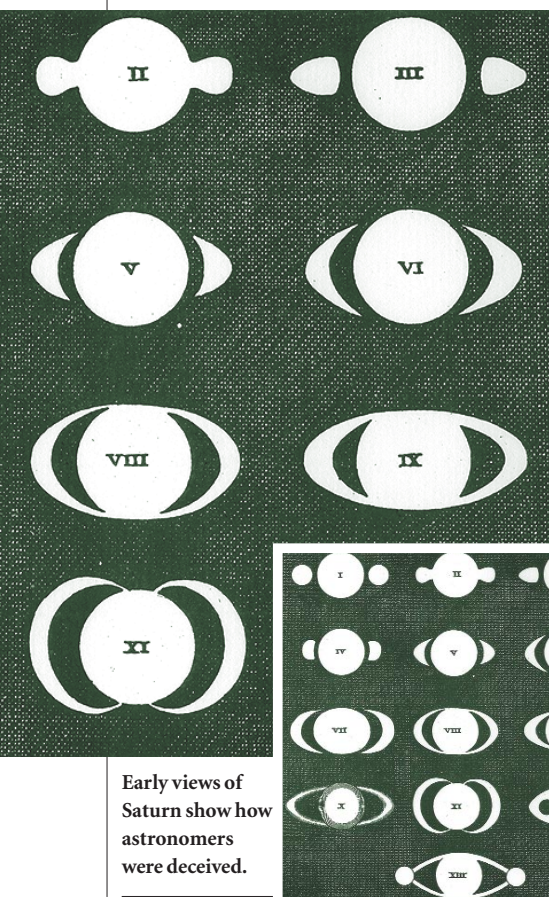
Laboratories of perception and illusion are familiar — as kitchens. A kitchen produces chemical reactions designed to evoke sensations and control consciousness in others as well as oneself. Measurements with kitchen instruments show a wealth of illusions. Small containers feel heavier than larger ones of the same-scale weight (with larger objects the muscles are set in expectation of a heavier weight). Some odd-shaped bottles appear to contain a greater volume. Colours affect taste. Strong tastes and smells adapt the tongue and nose so sensations change. Isn't the taste of wine affected by price? These everyday phenomena add greatly to our perceptions, though they are but illusions.

Science studies perception and illusion with methods of psycho-physics, which are more rigorous, but less varied, than cooking. Although lacking concepts of science, the cook has intimate knowledge of puzzles of mind and matter — mysteries lying in the hyphen between psycho and physics — where illusion reigns supreme. ■

Richard L. Gregory, Emeritus Professor of Neuropsychology, Department of Experimental Psychology, University of Bristol, Bristol BS8 1TN, UK.

FURTHER READING

Gregory, R. L. *Eye and Brain: The Psychology of Seeing* 5th edn (Oxford Univ. Press, 1997).
Gombrich, E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Phaidon, London, 1960).



Early views of Saturn show how astronomers were deceived.

Genie in a bottle

Robert J. Cava

An overlooked compound has a surprise in store for physicists. It becomes superconducting at a much higher temperature than any other stable metallic compound.

The field of superconductivity has been rocked by a startling announcement. For fifteen years, researchers have been delving into the mysterious and complex world of high-temperature superconducting materials — virtually ignoring simple metallic compounds because they superconduct at very low temperatures. But now Akimitsu and colleagues have discovered superconductivity at an amazing 39 degrees above absolute zero in the simple compound magnesium boride (MgB_2). They report their discovery on page 63 of this issue¹, in what must be one of the shortest communications published in *Nature* in recent memory.

Superconductors are materials that lose their resistance to electrical current flow below a certain critical temperature (T_c). In the ideal case, this zero-resistance state is absolute — electrons flowing in a continuous loop of superconducting wire below T_c could theoretically flow for the age of the Universe and never lose any energy. But in the real world there are losses from microscopic inhomogeneities, for example, and the ideal is never truly obtained.

Nonetheless, devices made with superconducting materials have resistances that are orders of magnitude lower than those of devices made with the best conventional conductors. This low resistance to current flow means that large currents (on the order of 10^6 amperes per square centimetre of wire cross-section) can be passed without significant heating. The magnets in magnetic resonance imaging instruments now in common use, for example, are made from metal-alloy superconducting wires. These magnets are cooled below the T_c of the metal alloy by immersion in liquid helium at 4.2 K. One can sometimes see trucks delivering helium to hospital loading docks for that purpose.

Almost exactly 15 years ago, physicists were stunned by the announcement that a ceramic composed of barium, yttrium, copper and oxygen could become superconducting at temperatures above that of liquid nitrogen (77 K)². This discovery, based on a modification of a formula first announced by Bednorz and Müller³ who later won the Nobel Prize for physics, sparked an explosion in condensed-matter physics and materials-science research, and the echo can still be heard. It is difficult to describe the feeling



Figure 1 The newly discovered superconductor magnesium boride has been available in large quantities from suppliers of inorganic chemicals for many years, but physicists have finally ‘rubbed the lamp’ and found that magnesium boride superconducts at an amazing 39 K (ref. 1).

that we had for the infinite possibilities promised by that discovery. Imagine a world with perpetual engines, trains that magnetically ‘float’ above the tracks and ultrafast computers. For the people in the thick of it, it was almost impossible to sleep for years afterwards — every minute spent sleeping was another minute missed in trying to figure out the implications of a whole new way of thinking about the world. Some of the promises of those early days have been fulfilled, and the legacy of the discovery of high-temperature superconductivity has been to change forever the culture of multidisciplinary research in the physical sciences.

Akimitsu apparently announced the discovery of superconductivity in MgB_2 (ref. 1) at a conference in Sendai, Japan, in early January. The story came to my attention a few weeks later, through what must have been a convoluted path of e-mails and word-of-mouth. The whole process is hauntingly reminiscent of the way such stories came to light in the early days of high-temperature superconductivity — under the guise of a narrative seemingly too fantastic to be true, and yet at the same time, too fantastic to be entirely false.

The story I heard was that Akimitsu and his group were attempting to make a chemical analogue of CaB_6 — a semiconducting material that surprisingly becomes ferromagnetic (like iron) when doped with a small amount of electrons⁴. They tried to

replace calcium with magnesium, which is directly above it in the periodic table. One of their starting materials was the simple compound MgB_2 , which has been known since 1953 and is available in kilogram-size bottles from suppliers of inorganic chemicals (Fig. 1). MgB_2 is one of the common reagents used in metathesis reactions (in which compounds exchange partners)⁵, and magnesium borides are used in some commercial preparations of elemental boron. Apparently, the stuff they got out of the bottle became superconducting at 39 K, 16 K higher than any other simple metallic compound. That must have been quite a shock.

But why the excitement? After all, critical temperatures for the high-temperature copper oxides have risen to 160 K over the years, four times the value for MgB_2 . There are two reasons for the fuss. First, early indications⁶ are that this material appears to become superconducting by what is known as the BCS mechanism (named after its discoverers, Bardeen, Cooper and Schrieffer)⁷, in which the interactions between the electrons that give rise to superconductivity are mediated by thermal vibrations of the atoms in the underlying crystal lattice. So, unlike the high-temperature copper oxide superconductors, MgB_2 is likely to be a ‘conventional’ superconductor — the rules of physics do not need to be bent for superconductivity to occur. MgB_2 has the highest T_c known for a chemically stable, bulk compound of this kind. This holds tremendous promise for

even higher superconducting temperatures in conventional materials.

Second, one of the as-yet unfulfilled promises of high- T_c copper oxide superconductors is the commercial production of wires carrying large amounts of current for everyday applications. There are several obstacles to making superconducting wires from copper oxides, in particular the passage of the superconducting current between the millions of tiny crystallites that make up the materials, although progress is continuously being made in this area. There is the chance that superconducting materials based on MgB_2 may eventually be able to carry more current than the copper oxide superconductors. With a T_c of 39 K, there is also the chance that they would not need to be cooled by liquid helium, but could be cooled by electrical refrigerators.

This discovery is the fulfillment of a promise made long ago for boride superconductors. Many predicted that the light mass of boron should lead to high lattice vibrational frequencies and therefore high superconducting transition temperatures (part of the BCS picture), but until now, although many boride superconductors have been discovered, boron's full potential in superconductivity has been unrealized.

If MgB_2 had been discovered in the 1960s and 1970s — at the time when such simple materials were being intensively tested for superconductivity — the whole culture of superconductivity research would have been

different. Some of the greats in those early days of discovery of simple compound superconductors worked on the metal diborides, yet they never tested this particular one. So for decades superconductivity was believed to be impossible for any compound above 30 K until the copper oxides proved otherwise, and was until recently believed by most to be impossible above 30 K for stable chemical compounds operating within the conventional BCS framework for superconductivity. I guess that story is different now.

If I didn't know better, I'd think we'd all been dreaming for the past few weeks. How much this discovery changes the path of materials physics depends on whether MgB_2 is a solitary example of a new way of making high-temperature superconductors or whether it represents only the tip of an iceberg. For the high- T_c copper oxides, we haven't found the bottom of that iceberg yet, even after 15 years of looking. I, for one, hope this iceberg is just as deep.

Robert J. Cava is in the Department of Chemistry and Materials Institute, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: rcava@princeton.edu

1. Nagamatsu, J., Nakagawa, N., Muraoka, T., Zenitani, Y. & Akimitsu, J. *Nature* **410**, 63–64 (2001).
2. Wu, M. K. *et al.* *Phys. Rev. Lett.* **58**, 908–911 (1987).
3. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–191 (1986).
4. Young, D. P. *et al.* *Nature* **397**, 412–414 (1999).
5. Killian, E. G. & Kaner, R. B. *Chem. Mater.* **8**, 333–343 (1996).
6. Bud'ko, S. L. *et al.* <http://xxx.lanl.gov/abs/cond-mat/0101463>
7. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175–1204 (1957).

attractant molecules enter the circulation, stimulating the migrating tumour cells to invade the walls of blood vessels and enter the organs.

There is evidence to support all three theories, but the mediator molecules — whether they be growth factors, adhesion molecules or chemoattractants — have remained unknown. Now, however, Müller *et al.*² have identified chemoattractants in target organs and chemoattractant receptors on tumour cells, providing molecular support for the chemoattraction theory. Moreover, the authors have shown that they can block metastasis in animals by blocking the receptors.

Circulating leukocytes (white blood

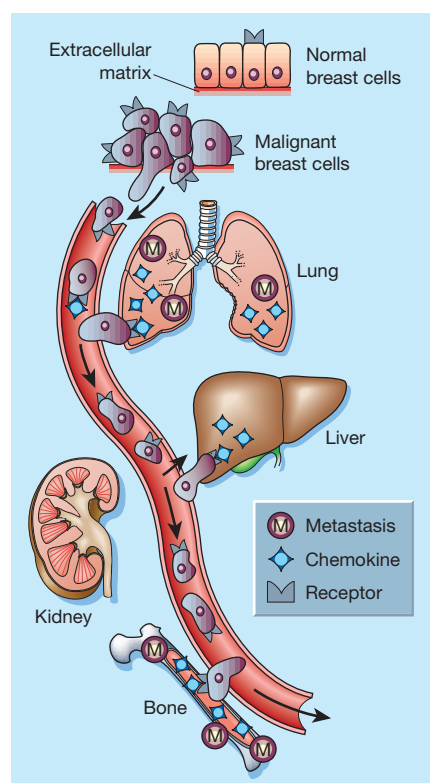


Figure 1 Targeted metastasis of primary breast cancer cells, according to the results of Müller *et al.*². Compared with normal breast epithelial cells (top), their malignant counterparts are enriched for the surface expression of the chemokine receptor CXCR4. The chemokines (such as CXCL12) that recognize these receptors are released in high quantities only by certain organs, such as bone marrow, liver, and lung. Other organs, such as kidney (pictured), skin and brain, contain low amounts of these chemokines. Malignant primary breast cancer cells invade their underlying extracellular matrix and circulate in the blood and lymphatic systems. Breast cancer cells passing through those organs with large amounts of the chemokines will be attracted to leave the circulation and enter the organs. So the final distribution of metastases reflects the relative abundance of chemokines in the different organs.

Cancer

An attractive force in metastasis

Lance A. Liotta

In breast-cancer patients, secondary tumours often form in the lungs and bone marrow, for example, but rarely in the kidneys. The explanation for this bias involves soluble attractant molecules called chemokines.

The main cause of treatment failure and death for cancer patients is metastasis — the formation of secondary tumours in organs a long way from the original cancer. As long ago as the early nineteenth century, it was recognized that secondary tumours are seeded by cells released from the original tumour and ferried about the body in the lymphatic and blood circulations. Autopsies reveal that some organs have large numbers of secondary tumours, while other organs have relatively few¹. But the basis for this bias has been a puzzle, because the pattern cannot be explained simply by differences in blood and lymph flow in different organs.

Writing on page 50 of this issue², Müller and colleagues provide a molecular answer to this old mystery. It turns out that soluble factors, released in large quantities from certain organs, can attract circulating cancer

cells to take up residence there. The authors identify the soluble factors to which human breast cancer cells are attracted, as well as the receptor proteins on the cancer cells that these factors engage.

Three major theories have been put forward to explain the bias of metastases (secondary tumours) towards certain organs^{3–6}, but studies of animals have failed to show which is right. The first theory proposes that tumour cells leave the blood and lymphatic systems to the same extent at all organs, but multiply only in those organs that have the appropriate growth factors. According to the second theory, the endothelial cells that line blood vessels in target organs express adhesion molecules that cause circulating tumour cells to put on their brakes and stop in those organs. Finally, the 'chemoattraction' theory holds that organ-specific

cells) and stem cells (undifferentiated cells with the potential to produce many types of more specialized cell) use chemoattractants called chemokines to home in on specific organs⁷. Chemokines are molecules that are structurally and functionally similar to growth factors. They bind to G-protein-coupled receptors on leukocytes and stem cells; the receptors are so-called because they work through guanine-nucleotide-binding (G) proteins to spark off intracellular signalling cascades that prompt migration towards the chemokine source. Chemokines induce leukocytes to adhere tightly to the endothelial cells that line the blood vessels and to migrate towards the highest concentration of chemokine. This migratory behaviour seemed identical to that of metastatic tumour cells, so Müller *et al.* proposed that tumour cells co-opt the same chemokines.

The authors conducted a comprehensive survey of chemokines and their receptors in tissue from breast-cancer patients. They found a receptor–chemokine pair (consisting of the receptor CXCR4 and the chemokine CXCL12) that fit the expected profile: CXCR4 was expressed more highly in breast-cancer tissue than in normal breast tissue; and CXCL12 was expressed in many organs, such as lymph nodes, bone marrow and lungs, in which breast cancer metastases are often found (Fig. 1). The authors also showed that *in vitro* the chemokine CXCL12 stimulated breast cancer cells to carry out the basics of invasion — the cells sent out extensions (pseudopodia), migrated in a directed manner, and penetrated barriers imposed by extracellular matrix.

Next, using immunodeficient mice, Müller and colleagues succeeded in blocking metastasis of injected human breast cancer cells to CXCL12-rich lung tissue by treating the mice with an antibody that neutralizes CXCR4 activity. The results led the authors to propose that small-molecule antagonists of chemokine receptors might one day be useful in treating cancer patients. As yet, though, it is unclear whether such treatment would have any effect on patients who already have established metastases.

Several other questions remain, too, such as why a breast cancer cell should behave like a leukocyte. Also, what is the selective pressure that causes the dominance of primary breast cancer cells that express high levels of CXCR4? It is presumed that malignant (invasive and metastatic) primary cancer cells are the outcome of a genetic progression process, by which a tumour cell acquires the mutations that allow it and its cellular offspring to become 'fitter' than normal and other tumour cells in a tissue, and hence predominant.

But why would the expression of a chemokine receptor that is important for homing to distant organs have survival value

at the initial tumour site? Indeed, what is the fitness test that selects out malignant primary cells? The fact that Müller *et al.* found a high level of CXCR4 in primary breast cancer cells, but not in normal breast tissue, indicates that the selection process is probably already complete before metastasis takes place. It will be important to see whether CXCR4 levels are high in premalignant breast cancer, which does not yet have the ability to invade or metastasize. Imagine that CXCR4 is not expressed even in premalignant cancers. This would implicate this receptor in the complex microenvironment that promotes the survival of primary tumour cells with invasive capacity. This

is speculative, of course, but if it is true the chemokine CXCL12 and its receptor CXCR4 might qualify as targets for 'chemoprevention' — a means of blocking the conversion of premalignant to invasive cancer. ■

Lance A. Liotta is in the Invasion and Metastasis Section, Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-1500, USA.
e-mail: lance@helix.nih.gov

1. Weiss, L. *Clin. Exp. Metastasis* **10**, 191–199 (1992).
2. Müller, A. *et al. Nature* **410**, 50–56 (2001).
3. Liotta, L. A. *et al. Cancer Res.* **36**, 889–894 (1976).
4. Fidler, I. J. & Hart, I. R. *Science* **217**, 998–1001 (1982).
5. Nicolson, G. L. *Bioessays* **13**, 337–342 (1991).
6. Tuck, A. B. *et al. Oncogene* **18**, 4237–4246 (1999).
7. Zlotnik, A. & Yoshie, O. *Immunity* **12**, 121–127 (2000).

Planetary science

Icing Ganymede

Louise M. Prockter

Much of Jupiter's moon Ganymede is covered in comparatively young ice. Images from spacecraft are providing clues about whether this resurfacing occurred primarily through tectonic or volcanic events.

Jupiter has 16 known satellites, one of which, Ganymede, is a whopper — it is the largest moon in the Solar System, and is bigger than Mercury. Its surface is divided into two very different types of terrain. About one-third of the moon is dark, heavily cratered and apparently ancient, and composed of a mixture of rock and ice. This dark terrain is sliced by swaths of bright, younger terrain of almost pure water-ice, which is either smooth or heavily fractured in texture¹ (Fig. 1a, overleaf). The bright terrain formed as Ganymede underwent some extreme resurfacing event, probably as the result of the moon's increase in size. This may have occurred as it entered a three-body resonance with siblings Europa and Io^{2,3}, causing internal melting.

On page 57 of this issue⁴, Schenk *et al.* add to the debate about how the bright terrain formed. Their work revives a model in which the terrain formed as the result of icy volcanism, in which water-ice magmas seeped up to the surface to fill low-lying areas. The resulting pattern seems to be a mixture of those of its nearest neighbours: Callisto has a dark, heavily cratered ancient surface, and Europa has a bright ice-covered surface that is sparsely cratered and extremely young. Studying the process that ripped apart two-thirds of Ganymede's surface, yet left the remainder virtually untouched, will help us to understand the evolution of Jupiter's moons.

Images from the spacecraft Voyager, obtained in 1979, led to the proposal that the bright terrain formed when faulted blocks dropped down to create shallow troughs, which were subsequently flooded by icy vol-

canic material, either liquid water or glacial ice⁵ (Fig. 1b). This model could account for the apparent smoothness of some parts of the surface; the fracturing observed in the rest was thought to result from faulting that occurred later. Some scalloped depressions, loosely resembling the calderas (craters) of terrestrial volcanoes, were thought to fit in with this cryovolcanic picture.

In the late 1990s the Galileo spacecraft delivered images with vastly enhanced resolution. Features that one would expect to provide evidence of icy volcanism — vents, flow lobes or the embayment patterns seen when a fluid runs into a depression — were conspicuous by their absence. But the Galileo images did show that Ganymede's bright terrain had experienced considerable tectonic activity, supporting the idea that severe faulting had occurred. This led to the proposal that the bright terrain might have been resurfaced primarily by tectonic activity^{6–8}. According to this model, fracturing and faulting mean that the previous surface is no longer recognizable, and features such as craters, which are diagnostic of age, have been completely erased (Fig. 1c). In this scheme, Ganymede's dark terrain consists of a layer of dirty, ice-poor material overlying a brighter icy subsurface⁹: tectonic activity displaces the dark material and reveals the bright, icy subsurface.

Schenk *et al.*⁴ have tackled the issue of Ganymede's bright terrain by resourceful use of both Voyager and Galileo data sets. They have combined images from both spacecraft to create stereo pairs — two images that have sufficient separation for topography to be derived. These stereo pairs

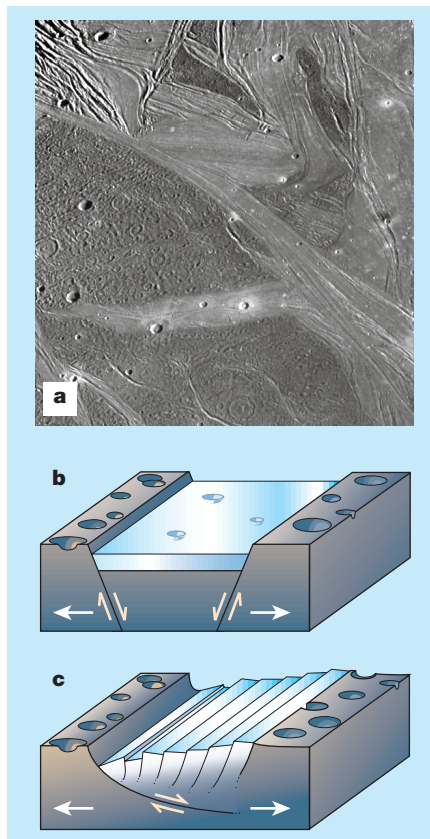


Figure 1 Resurfacing Ganymede. a, An image taken by the Galileo spacecraft showing the dark ancient terrain and the bright resurfaced areas. b, A model, based on data from the spacecraft Voyager, in which the bright terrain formed by initial faulting of surface rock into blocks. These fractured blocks dropped down to create lower-lying regions, which were flooded by water-ice magmas seeping to the surface by cryovolcanism³. c, A Galileo-based model in which the bright regions formed through the disruption of dark terrain by tectonic activity. That activity caused loose surface material to fall down hill, revealing the brighter, ice-rich substrate. Schenk *et al.*⁴ conclude that both volcanic and tectonic processes were involved, but that they operated at different heights.

allow the authors to identify vertical variations of about 200 m on Ganymede's surface. Topography can be a useful tool for finding out how and when surfaces formed and deformed.

By studying several areas on Ganymede, the authors find a number of smooth swaths of bright material that are also the lowest topographically; the most heavily fractured material tends to be much higher. Their findings are consistent with cross-cutting relationships among bright swaths, in which older segments of bright terrain are over-printed or destroyed by younger swaths. The authors propose that the smoothest areas of bright terrain are the result of resurfacing through icy volcanism, whereas the higher-standing, heavily fractured regions were resurfaced tectonically.

The difficulty with the notion of icy volcanism is that negatively buoyant liquid-water magma must have come up through a solid, ice-rich crust. By showing that the smooth swaths are up to a kilometre lower than their surroundings, Schenk *et al.* propose that this is the height to which liquid water can rise up through Ganymede's crust before it stops at the level of neutral buoyancy. They suggest that the higher topography was heavily grooved by faulting and tectonic deformation, but was too high for water-ice magma to reach and flood its surface.

These findings draw together two differing schools of thought. But although Schenk and colleagues' models show that the smoothest bright swaths are low-lying, there

is little evidence that they were subsequently flooded with water-ice magma. If such a plentiful source of icy magma exists at shallow depth, why is it so shy that it fails to show itself? And what is the relationship, if any, between the caldera-like features and the smooth swaths? Only one such feature shows associated flow-like patterns in the Galileo images¹⁰.

It was only quite recently, in May 2000, that Galileo took its highest-resolution images of Ganymede, so they have not yet been fully analysed. They should tell us more about the relative effects of volcanic and tectonic activity on this giant among moons, and so about the evolution of both its surface and its interior.

Louise M. Prockter is in the Space Physics Group, Applied Physics Laboratory, The Johns Hopkins University MS 7-366, 11100 Johns Hopkins Road, Laurel, Maryland 20723-6099, USA.
e-mail: louise.prockter@jhuapl.edu

1. Shoemaker, E. M., Lucchitta, B. K., Wilhelms D. E., Plescia, J. B. & Squyres, S. W. in *Satellites of Jupiter* (ed. Morrison, D.) 435–520 (Univ. Arizona Press, Tucson, 1982).
2. McKinnon, W. B. *Proc. Lunar Planet Sci. Conf.* 12, 1585–1597 (1981).
3. Showman, A. P. & Malhotra, R. *Icarus* 127, 93–111 (1997).
4. Schenk, P. M., McKinnon, W. B., Gwynn, D. & Moore, J. M. *Nature* 410, 57–60 (2001).
5. Parmentier, E. M., Squyres, S. W., Head, J. W. & Allison, M. L. *Nature* 295, 290–293 (1982).
6. Head, J. W., Pappalardo, R. T., Collins, G. & Greeley, R. *Proc. Lunar Planet Sci. Conf. XXIX* [CD-Rom] abstr. 535 (1998).
7. Pappalardo, R. T. *et al. Icarus* 135, 276–301 (1998).
8. Collins, G. C., Head, J. W. & Pappalardo, R. T. *Icarus* 135, 345–359 (1998).
9. Prockter, L. *et al. Icarus* 135, 317–344 (1998).
10. Kay, J. E. & Head, J. W. *Proc. Lunar Planet Sci. Conf. XXX* [CD-Rom] abstr. 1103 (1999).



100 YEARS AGO

A circular letter has been sent out seeking an expression of opinion from experts as to the advisability of founding a journal for the statistical study of biological problems... *Biometrika* is the proposed title of the journal; thirty shillings the estimated cost of the first volume. Statistical work in biology, to be of service, must be far-reaching and extensive, and it cannot be dissociated from morphological enquiry of the better kind. Progress must necessarily be slow, and the accumulation of results worth publishing only be expected after protracted research; and in these circumstances we are doubtful if the desire to burden the already overcrowded literature of biology with a new serial is not somewhat premature. It might be borne in mind that existing periodicals and the organs of societies are available for purposes of publication; and we could well desire for some of these that much of the so-called "systematic" work and quibbling over priority in nomenclature, fast becoming intolerable, might be replaced in work of the statistical and experimental order.
From *Nature* 28 February 1901.

50 YEARS AGO

The successful use of antibiotic substances against human disease is one of the most fascinating developments of modern medical science. Seven such substances are now commercially available, and the one recent introduction — terramycin — formed the subject of a recent conference under the auspices of the New York Academy of Sciences. A considerable part of the report collects information on the action of terramycin against human diseases... Those diseases in which positive action is recorded include pneumococcal pneumonia, rickettsial diseases, brucellosis and some venereal diseases. Terramycin appears to be ineffective against infections of *Pseudomonas aeruginosa* and *Bacillus proteus*; and of virus diseases, mumps, measles, chicken-pox and smallpox do not respond to treatment. Terramycin is only a very recent discovery from the biochemical research laboratories of Chas. Pfizer and Co., Inc. It was found as a result of a carefully planned research programme involving the examination of micro-organisms in soil from various parts of the world. The present imposing report of available information reflects a very speedy reaction to its implications within the medical profession.
From *Nature* 3 March 1951.

Particle physics

Precision precession

Frank Wilczek

The most accurate measurement yet of the way an elementary particle wobbles — precesses — in a magnetic field is getting physicists excited. If it is right, we may be on the threshold of a new era of particle discoveries.

A large collaboration of scientists, known as experiment E821 at Brookhaven National Laboratory, announced on 9 February the latest result from their study of the behaviour of fundamental particles (specifically muons) in a magnetic field¹. Their measurement of the magnetic moment of the positive muon has an uncertainty of just over one part per billion, three times better than the existing value. Moreover, it deviates in a small but important way from the prediction of the theory of matter (also known as the standard model — a woefully inadequate description). At this level of precision, the behaviour of the muon becomes sensitive to the existence of new heavy particles, such as those required by supersymmetry, a very popular but still hypothetical extension of the standard model. Although not yet statistically airtight, the E821 result could be a glimpse into a new world of physical phenomena.

The story of magnetic moments is brief, but glorious. Evidence that the electron had a magnetic moment was first gleaned in 1925 by two graduate students, Samuel Goudsmit and George Uhlenbeck. By carefully studying atomic energy levels (spectra), they discovered discrepancies with prevailing theories that could be resolved if it were assumed that all electrons act as tiny bar magnets. The strength of these elementary magnets is commonly given in terms of the gyromagnetic ratio, g , the relationship between the magnetization of the particle and its rotation. In a magnetic field the axis of a particle's intrinsic rotation, or spin, precesses in the same way that a toy top wobbles as it spins. The frequency of the particle's precession is μB , where B is the size of the magnetic field and μ is its magnetic moment given by gsq/mc , where s , q , m are the spin, charge and mass of the particle, respectively, and c is the speed of light.

Classical physics predicts $g = 1$ for a spinning charged particle, but to fit their observations Goudsmit and Uhlenbeck required $g \sim 2$ (where g is g for the electron). In 1928 Paul Dirac proposed a modification of the Schrödinger equation describing the quantum behaviour of electrons, to make it consistent with the special theory of relativity. Though it was motivated from unrelated considerations, Dirac's equation predicts $g = 2$ for simple, point-like electrons, providing a neat agreement between experiment and theory that lasted two decades. Then, in 1947, Polykarp Kusch and collaborators

determined that the experimental value of g is actually slightly bigger than Dirac's value, by approximately $g_e - 2 = 0.00236$.

Theorists quickly appreciated the significance of this discrepancy. The Dirac theory applies to an idealized electron, existing in isolation in empty space. But real observable electrons are surrounded by virtual particles that briefly pop in and out of existence in the quantum 'vacuum'. In one sense virtual particles are purely theoretical constructs, because they are by definition too short-lived to be observed directly, but to physicists they are quite real and have a tangible influence on the particles we observe.

In 1948 Julian Schwinger calculated that the interaction of electrons with virtual photons modifies their magnetic moment in a way that quantitatively explains Kusch's result. Richard Feynman did the calculation in a different and simpler way that could be generalized. Using these techniques, and the modern theory of matter, magnetic moments can be calculated with extraordinary precision. For this work, Dirac won the Nobel Prize for physics in 1933, Kusch in 1955, and Schwinger, Tomonaga and Feynman in 1965.

The same calculations hold for muons, which are elementary particles that resemble electrons in every way except that they are about 200 times as heavy. They have a half-life of about a microsecond, and so are not found on Earth in ordinary matter, although they do appear in cosmic rays. They are more con-

venient to study than electrons because the muon's magnetic moment is considerably more sensitive to the effects of heavy particles. The strength of the muon's bar magnet, g_μ , can be measured in a high-quality magnetic field by monitoring how rapidly it precesses. But to reach the precision of the Brookhaven experiment requires the world's largest superconducting magnet (Fig. 1) and measurements from billions of muons. The experiment is a *tour de force* of modern experimental technique, despite being relatively cheap by high-energy physics standards. The character of the result from E821 is best conveyed through its accurate quotation: $g_\mu - 2 = 0.0023318319 (\pm 0.0000000013)$.

Theorists have also worked hard to reach this level of accuracy. Their calculations must include not only the effects of virtual photons, but also more complex effects that arise because the properties of the virtual photons are themselves modified by interactions with virtual electrons, muons, tau leptons and quarks. Moreover, the small effects of virtual versions of much heavier particles, such as W and Z bosons, come into play. When all this is taken into account, there is a mismatch between theory and experiment of about four parts per billion. The discrepancy is about 2.6 standard deviations — highly suggestive, but not yet definitive.

When assessing loose talk about shaking the foundations of physics or overthrowing the standard model, which has already appeared in the popular press, we should keep in mind the size of the reported effect. Ordinarily, a theory whose rigorous application predicts the results of subtle experimental results with part-per-billion accuracy might be considered to be doing pretty well! So it is here. For this reason, the most plausible explanations of the discrepancy — barring computational or experimental error, or a statistical fluke — involve adding to the accepted theory



Figure 1 The 14-metre-diameter superconducting magnet used by the E821 experiment at Brookhaven National Laboratory, Long Island, New York. It provides a magnetic field of extremely high quality, allowing physicists to measure the magnetic behaviour of the muon with unprecedented accuracy.

BROOKHAVEN

of matter, rather than overthrowing it. Indeed, the simplest explanation is that there is some new species of virtual particle that has inadvertently been left out of the theoretical calculation, because its analogue in the real world is very heavy and unstable and has so far escaped detection in particle accelerators.

Particularly attractive, because it is also suggested by several other pieces of indirect evidence, is the possibility that some of these particles are supersymmetric². Supersymmetry theory predicts that every fundamental particle in the standard model has a heavier superpartner — the selectron for the electron and smuon for the muon. If supersymmetry is true, it provides new virtual particles galore, with masses of 100–1,000 times the mass of a proton. Our knowledge of the correct model of supersymmetry to use — if any — is limited, so no precise prediction is possible. But the observed E821 discrepancy

sits quite comfortably in the range preferred by models of low-energy supersymmetry³.

The reported E821 results are based on analysis of data taken in 1999. A data sample from 2000, containing four times as many positive muons, awaits analysis. Similar data are now being taken for negative muons, which should be able to confirm whether particles and antiparticles behave symmetrically. Elsewhere, work currently underway will narrow down residual uncertainties in the predictions of the standard model. A year from now we'll either be celebrating at the oasis — or mourning the mirage.

Frank Wilczek is at the Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.
e-mail: wilczek@mit.edu

1. Brown, H. *et al.* <http://xxx.lanl.gov/abs/hep-ex/0102017>
2. Wilczek, F. <http://xxx.lanl.gov/abs/hep-ph/0101187>
3. Czarnecki, A. & Marciano, W. <http://xxx.lanl.gov/abs/hep-ph/0010194>

Molecular biology

A big development for a small RNA

David A. Clayton

The RNA-processing enzyme MRP contains an RNA component that is essential for its activity. Unexpectedly, it seems that mutations in the gene encoding this RNA cause a multifaceted human disease.

Over 35 years ago, some members of the Amish families in Pennsylvania were found to be suffering from an unusual inherited disease. Later, the disorder was also discovered in a small fraction of the Finnish population. People with the disease are unusually short, have changes in the structure and abundance of their hair, and show abnormalities in the development and function of their bone and cartilage, so the disease was given the name cartilage–hair hypoplasia (CHH)¹. Affected people may also have a compromised immune system, making them more susceptible to infection and predisposing them to cancer, particularly tumours of the lymphatic organs.

Which gene, or genes, might be mutated in people with CHH? In earlier genetic studies, the culprit was tracked down to one of the non-sex chromosomes, and was mapped to a region on chromosome 9 (ref. 2). It was also discovered that the disorder is a recessive condition¹, meaning that, for the disease to occur, both copies of chromosome 9 — one from each parent — must bear mutations in that chromosomal region. Yet, despite these crucial discoveries, attempts to pinpoint the underlying gene have been unsuccessful — until now. The answer, revealed by Ridanpää and colleagues³ in a recent issue of *Cell*, is unexpected.

In refining the map position of the chromosomal region that is associated with CHH, Ridanpää *et al.*³ disclosed 11 candidate genes

that might underlie the disease. Ten of these, as expected, encode proteins. But, when the authors examined these genes from CHH patients, they did not find any mutations that were likely to have caused the disease. The eleventh gene from these patients, however, did have potentially relevant mutations. This gene is different because it is the only known nuclear gene that encodes the RNA component of an RNA-cleaving enzyme, a ribonuclease known as MRP. The RNA encoded by the gene is an essential part of the enzyme, but it is not translated into protein.

This unanticipated finding, that alterations in the gene encoding the RNA component of an RNA-processing enzyme are at least partly responsible for the complex physical hallmarks of CHH, immediately shifts our attention to the enzyme itself (Fig. 1). MRP is a large protein–RNA complex that was first characterized for its ability to cleave the 'primer' RNAs needed for copying mitochondrial DNA⁴ — hence the name MRP, for 'mitochondrial RNA processing'. But most MRP is localized to a subsection of the nucleus (the nucleolus), where it participates in a late stage in the processing of the so-called 5.8S ribosomal RNA in yeast, and presumably in other organisms as well⁵.

The CHH-associated mutations that Ridanpää *et al.* found in the gene encoding MRP's RNA component were insertions or duplications of sequence in the promoter (a

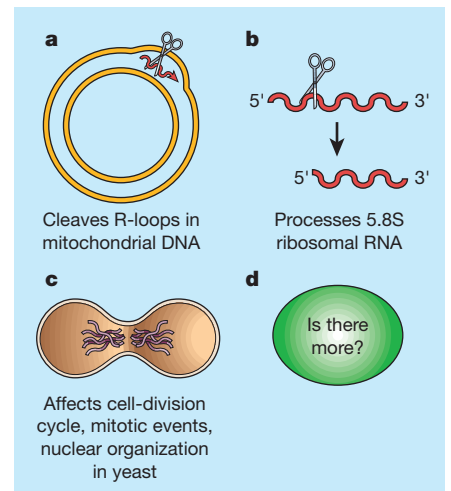


Figure 1 Functions of the RNA-cleaving enzyme (ribonuclease) MRP. Ridanpää *et al.*³ show that mutations in the gene encoding the RNA component of MRP underlie an inherited disorder, cartilage–hair hypoplasia. These mutations could affect several cellular processes. a, Processing of mitochondrial 'primer' RNA (wavy line) during the replication of mitochondrial DNA. The mitochondrial DNA is shown as a double ring. The primer RNA must first associate with the DNA template, forming an RNA–DNA hybrid. The MRP (shown as scissors) then cleaves the primer RNA, leaving it with free 3' ends, to which bases can be added by the mitochondrial DNA-copying machinery (DNA polymerase; not shown). Loss of MRP activity in mitochondria might lead to a depletion of mitochondrial DNA. b, A late step in the processing of precursor RNA to mature forms of 5.8S ribosomal RNA. The ribosome is required for protein translation, so loss of MRP activity might hamper translation of key proteins. c, In yeasts, mutations in the RNA¹¹ or protein¹⁰ components of the MRP result in alterations in the mitotic cell-division cycle and cell morphology. The exact role of MRP RNA here is unknown. d, MRP might have yet more functions, either as a site-specific ribonuclease, or together with other activities.

control region), or principally single base changes in the RNA-coding part. The authors looked at cells from two patients who had both types of mutation: one copy of the gene had mutations only in the promoter region; the other copy had mutations only in the coding sequence. These cells contained less than half the amount of the MRP RNA found in cells from controls. And this RNA came only from the genes that had mutations in the coding sequence, implying that the reduced RNA levels were caused by alterations in the structure of the promoters of the other genes, and hence in the control of gene expression. Cells from patients with mutations only in the RNA-coding part expressed the RNA at normal levels. Interestingly, these coding-sequence mutations

reside in regions that have been highly preserved during evolution, and are reminiscent of those that generate altered physical characteristics in yeast⁶. In fact, the RNA is thought to be essential in yeast⁷. Ridanpää *et al.* discovered no patients in whom the relevant gene was absent, or in whom both copies of the gene were not expressed at all, implying that the RNA is equally important in humans.

Are there any clues in all of this to how mutations in the MRP RNA component might produce the varied physical characteristics seen in CHH? Given that MRP is involved in copying mitochondrial DNA, might some dysfunction in mitochondria be responsible for CHH? This seems unlikely, for at least two reasons. First, although some of the physical features of people with CHH overlap with those of people with mitochondrial disorders, most do not. Second, the vast majority of MRP is found in the nucleolus. The reduced levels of the MRP RNA in patients with CHH are still greatly in excess of those needed for mitochondria to function properly.

But it is less clear whether processing of the 5.8S ribosomal RNA in the nucleolus would be problematic. The ribosome is needed to translate messenger RNAs into proteins, so the correct processing of the 5.8S RNA is crucial for translation. It seems reasonable, then, that the mutations seen in CHH might hamper the accurate processing of the 5.8S ribosomal RNA at times in development when key proteins need to be synthesized.

Another possibility is that we need to look to other pathways, and other organisms, for hints as to how these mutations in the MRP RNA gene produce functional defects. There is no clear precedent for a nuclear, RNA-encoding gene causing a major disease, but this could be because — unlike the MRP RNA gene — there are generally multiple copies of such genes in the human genome. Even so, mutations in multi-copy human mitochondrial DNAs encoding transfer and ribosomal RNAs are involved in mitochondrial diseases⁸, with mouse models now emerging for studying the nature and progression of a given disorder⁹.

An intriguing possibility comes from a recent study of yeast (cited by Ridanpää *et al.*), in which a mutation in a protein component of MRP leads to a prolonged and abnormal arrest in cell division¹⁰. Similar defects in people with CHH would be consistent with their failure to develop a normal number of immune cells, for example. Moreover, mutations in the RNA component of MRP in the fission yeast *Schizosaccharomyces pombe* result in atypical cellular morphology¹¹. We should not be surprised if this RNA has even more functions. But, given that the mutations within the coding part of the gene are found in both CHH patients and

yeast, and that these mutations are likely to affect the function of the MRP enzyme, it would be surprising indeed if RNA processing were not part of the final picture of cartilage–hair hypoplasia.

David A. Clayton is at the Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, USA.

e-mail: clayton@hhmi.org

1. McKusick, V. A., Eldridge, R., Hostettler, J. A., Ruangwit, U. & Egeland, J. A. *Bull. Johns Hopkins Hosp.* 116, 231–272 (1965).

2. Sulisano, T. *et al. Nature Genet.* 3, 338–341 (1993).
3. Ridanpää, M. *et al. Cell* 104, 195–203 (2001).
4. Chang, D. D. & Clayton, D. A. *Science* 235, 1178–1184 (1987).
5. Clayton, D. A. *Proc. Natl Acad. Sci. USA* 91, 4615–4617 (1994).
6. Shadel, G. S., Buckenmeyer, G. A., Clayton, D. A. & Schmitt, M. E. *Gene* 245, 175–184 (2000).
7. Schmitt, M. E. & Clayton, D. A. *Genes Dev.* 6, 1975–1985 (1992).
8. Wallace, D. C. *Science* 283, 1482–1488 (1999).
9. Johnson, K. R., Zheng, Q. Y., Bykhovskaya, Y., Spirina, O. & Fischel-Ghodsian, N. *Nature Genet.* 27, 191–194 (2001).
10. Cai, T., Reilly, T. R., Cerio, M. & Schmitt, M. E. *Mol. Cell. Biol.* 19, 7857–7869 (1999).
11. Paluh, J. L. & Clayton, D. A. *EMBO J.* 15, 4723–4733 (1996).

Animal husbandry

Assessing the welfare state

Michael Mendl

Animals kept for human use should surely have conditions that are as conducive to their welfare as possible. A study of captive mink illustrates how animal welfare can be assessed.

Do mink in fur farms suffer if they are denied access to resources such as water for swimming? This seemingly esoteric question encapsulates much of the debate about animal welfare and the conditions in which animals are kept on farms, in laboratories and in zoos. It is also deceptively complicated, as is evident from the item by Mason and colleagues on page 35 of this issue¹. The take-home message of their paper, however, is that science can inform argument about animal welfare by providing objective evidence about how animals respond to captivity.

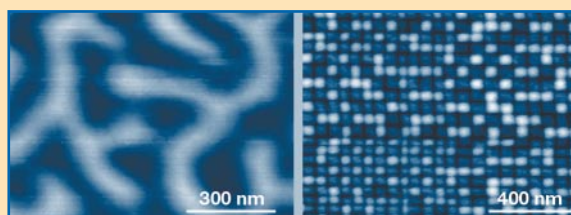
There are two main scientific approaches to the issues². One is based on the proposal that animal welfare is compromised when normal biological function is impaired^{3,4} — as seen, for instance, in low growth rates, behavioural abnormalities and immunosuppression. The aim of this approach is to measure a variety of 'welfare indicators' to obtain a comprehensive view of the animal's biological state⁴. But interpreting and integrating the measures to provide an overall welfare assessment is by no means straightforward^{2,4}. For instance, is an animal that exhibits high levels of abnormal behaviour

Materials science

All chopped up

In the constant push for higher-capacity hard drives, K. Koike *et al.* have found a way of doubling the density of data storage in magnetic media (*Applied Physics Letters* 78, 784–786; 2001).

Conventional high-density magnetic recording media consist of films composed of magnetic grains. Bits of data are stored on these grains, the size of which limits the amount of stored data. More data can be encoded if the grains are smaller and more tightly packed. But if the grains are too small, they become unstable and can 'forget' their information because the domains (magnetized regions) produced by the grains can affect each other through thermal instability. In other words, a '1' can be changed



into a '0' by a small temperature fluctuation. The highest density of reliable data storage on so-called continuous magnetic media reported so far is 65 billion dots per square inch. An image of continuous media showing the domains is seen above on the left.

Koike *et al.* chopped up a magnetic block to double the density of storage, with potential to increase this even further. They reduce the problem of thermal instability by physically separating clusters of grains

using focused ion beam lithography (an atomic knife) to cut grooves into the block. The image on the right shows domains from these discontinuous media: this denser pattern can store more data.

So can we expect colossal hard drives soon? Not just yet. The fabrication process has not been perfected, and data recording and retrieval are very slow. But if these problems are resolved, we may have even more games, movies and music at our fingertips. **Josette Chen**

but low physiological stress responses better or worse off than one displaying the reverse?

The second approach equates an animal's welfare with its subjective experiences^{5,6}. Animals may grow, reproduce and appear to be healthy, yet still have poor welfare if they experience subjective suffering such as prolonged frustration from having little space in which to move. Here, largely, is the cause of public concern about animal welfare — if non-human animals were shown to lack conscious subjective experiences, that concern would almost certainly wane. For the scientist, the problem is that subjective states such as those that we call frustration, anxiety or pain are difficult to measure or cannot be measured at all.

Advocates of this approach have therefore developed methods based on the assumption that animals suffer if denied resources that they are strongly motivated to obtain, or exposed to stimuli that they will work hard to avoid⁵. Theories that emotions are important in motivating animal behaviour lend support to this assumption^{7,8}. The methods involve finding out how much animals value resources by testing their preferences or, using techniques borrowed from human consumer economics, by measuring how high a price, in terms of time and energy, they will pay to access or avoid them^{5,9}. These 'motivation'-based methods have limitations⁵ — for example, sweet foods may be preferred in the short term but do not enhance health and welfare in the long term — but they circumvent the problem of integrating many different welfare indicators.

Most people in this field acknowledge that the 'welfare indicators' and 'motivation' methods complement each other^{2,4,5}. Even so, the two approaches have largely been used in isolation — Mason and colleagues¹ achievement is to have brought them together. The authors investigated whether the small wire-mesh cages used to house farmed mink cause frustration by preventing natural activities such as swimming or using several nest sites. Sixteen individually housed mink were each given access to seven cages containing resources such as an alternative nest site or a water pool. When made to work for access to the resources by pushing through weighted doors, mink worked hardest for the water pool. This conclusion was reached using four different calculations of how the mink valued each resource — a comprehensive analysis that sets new standards in animal welfare studies.

Following the assumption of the 'motivation' approach, these findings suggest that depriving mink of a water pool may result in suffering and poor welfare. Mason *et al.* tested this assumption by measuring changes in welfare indicators when the mink were prevented from visiting the pool for 24 hours. They observed a rise in urinary cortisol, an endocrine indicator of stress, that was not

accompanied by an increase in activity and was indistinguishable from that seen when an essential resource, food, was removed for the same duration.

By using the 'motivation' and 'welfare' methodologies together, Mason *et al.* provide powerful evidence that the mink value the water pool highly and exhibit stress responses in its absence. So lack of a water pool may compromise the welfare of mink even if their growth and health are good. Answers to a number of remaining questions could strengthen this conclusion. The roles of expectation and experience of a resource in determining response to deprivation remain to be tackled. So too does the problem that resource valuation depends on what else is available (for instance, chewable objects might have been valued more highly if they were the only resource offered). Furthermore, deprivation of a valued resource may be stressful only in the short term, or if no adequate substitutes are available. For example, removing the highly valued alternative nest site did not induce a cortisol rise, perhaps because the animal's own nest was still accessible.

What of the wider implications? Captive animals are often denied resources or opportunities to behave in ways that, intuitively, would seem to be important to them. Confinement prevents domestic sows from building nests before they give birth; there

is usually no dust-bathing substrate in cages for laying hens; and most laboratory rodents have limited opportunity to construct burrows or enclosed nests. Techniques such as those used by Mason *et al.* should allow us to go a good deal further than intuition in deciding whether these resources and opportunities are indeed highly valued by the animals, and whether their absence causes frustration and stress. If so, scientific argument for their provision will be strong and should form the basis for decisions to alter housing conditions, even though such decisions will inevitably be coloured by political, economic and practical considerations. ■

Michael Mendl is in the Centre for Behavioural Biology, Department of Clinical Veterinary Science, University of Bristol, Langford House, Langford BS40 5DU, UK.

e-mail: mike.mendl@bris.ac.uk

1. Mason, G. J., Cooper, J. & Clarebrough, C. *Nature* **410**, 35–36 (2001).
2. Duncan, I. J. H. & Fraser, D. in *Animal Welfare* (eds Appleby, M. C. & Hughes, B. O.) 19–31 (CAB Int., Wallingford, Oxon, 1997).
3. McGlone, J. J. *J. Agric. Environ. Ethics* **6** (suppl. 2), 26–36 (1993).
4. Broom, D. M. & Johnson, K. G. *Stress and Animal Welfare* (Chapman & Hall, London, 1993).
5. Dawkins, M. S. *Behav. Brain. Sci.* **13**, 1–61 (1990).
6. Duncan, I. J. H. & Petherick, J. C. *J. Anim. Sci.* **69**, 5017–5022 (1991).
7. Cabanac, M. *J. Theor. Biol.* **155**, 173–200 (1992).
8. Rolls, E. T. *Behav. Brain. Sci.* **23**, 177–234 (2000).
9. Houston, A. I. *Anim. Behav.* **53**, 983–990 (1997).

Condensed matter

Memories of paste

David A. Weitz

Pastes are not the simple materials they appear to be. It seems they have a 'memory': after a force has been applied, they recover and move back in the opposite direction.

Pastes are familiar materials we use every day, for example, when we brush our teeth. They are prototypic 'soft' materials: they behave as a solid until a sufficiently large load or stress is applied, at which point they flow like a fluid. So a squeeze is all that is required to put toothpaste on a brush. But the way pastes and similar soft solids behave remains poorly understood, partly because it is difficult to perform reproducible experiments on these materials — their properties seem to change with time and depend strongly on their history. As they discuss in *Physical Review Letters*, however, Cloitre, Borrega and Leibler² have developed new experimental techniques and show that the way a paste recovers from an applied stress is remarkably like the behaviour of glasses and gels^{3,4}.

Pastes typically consist of a suspension of small particles in a background fluid. These particles are crowded, or jammed together

like grains of sand on a beach, forming a disordered, glassy or amorphous structure, and giving pastes their solid-like character. This jamming together gives pastes some of their most unusual properties. The particles exert large forces on their neighbours, but in solid pastes the material, and hence each individual particle, is stationary, so the net sum of the forces on each particle must be zero. Because of the high degree of disorder inherent in any paste, some of these local force balances will be precarious. A very small change, perhaps induced by a fluctuation in temperature, can disrupt the balance, causing a relaxation in the structure that can ripple through a large volume because of the close proximity of the particles and the requirement that all local forces have to be balanced.

After relaxation from such a small disturbance, the new state is likely to be more stable and disruptions will be less probable.

A slightly greater disruption is required to cause further relaxation. This results in the paste 'ageing' — it relaxes more slowly as the waiting time increases or as the sample gets 'older'. The relaxation continues to slow as time goes by, so the ageing process never comes to a complete stop. The time that a relaxation takes to occur typically depends on the paste's age, and can be plotted onto a logarithmic curve, known as a master curve.

In the pastes studied by Cloitre *et al.*², the particles are microgels, which are small chunks of gel swollen by a solvent. At low concentrations, these microgels are not jammed together, and their properties are determined by the solvent. But as their number is increased, the microgel particles begin to squeeze on each other, creating the paste-like behaviour.

Rather than study the structure of the pastes directly, Cloitre *et al.* instead look at their rheological, or mechanical, properties, which determine both flow and structure. Because the relaxation in the solid paste depends on its age, the authors set up an initial state as a reference point from which to determine the age of the sample. They do this by applying a very large shear stress which 'melts' the sample, changing the paste to a fluid-like state. The ageing begins when the stress is removed and the paste becomes a solid again. Cloitre *et al.* then probe the mechanical properties of the paste by using very small stresses that are just enough to elicit a measurable response.

The authors make a remarkable observation: although the sample was completely fluidized by the large shear stress, it developed a 'memory' of the direction in which the stress was applied, and the solid-like paste slowly 'pulled back' on itself in the opposite direction, eventually passing beyond its initial position. This 'recovery' follows the logarithmic master curve that typifies the ageing behaviour. The memory effect is robust. The direction in which the small additional stress is applied does not matter: the long-time strain recovery is always in the opposite direction.

Similar ageing effects are observed in measurements of long-term 'creep'; here a stress that is just enough to make the sample flow is applied some time after the sample is melted. The flow or creep does not begin immediately, but instead requires some onset time, which again depends on the waiting time. Similar effects are observed in the ageing of glasses, and Cloitre *et al.* use an analysis akin to that used in the description of glassy polymers: like the paste samples of Cloitre *et al.*, such glasses display ageing, and the relaxation data can again be scaled onto a single master curve.

What causes this ageing behaviour and the memory effect? The microscopic origin

is unclear — local forces may be involved, but why they lead to a memory and a strain in the negative direction is not clear. Nevertheless, the great appeal of these simple pictures for ageing is that, even without complete understanding, it is still possible to scale the data onto universal curves using the age of the sample as a scaling parameter. The question that must now be addressed is how much impact this scaling behaviour, the ageing and the highly unusual rheology have. Are they fascinating peculiarities of this system, depending on the microscopic nature of the microgels? Or are they symptomatic of more general behaviour that will be observed

in other materials? If the latter, this may have implications for manufacturing and processing of soft materials in industrial applications. ■

David A. Weitz is in the Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138 USA.

e-mail: weitz@deas.harvard.edu

1. Sollich, P., Lequeux, E., Hébraud, P. & Cates, M. E. *Phys. Rev. Lett.* **78**, 2020–2023 (1997).
2. Cloitre, M., Borrega, R. & Leibler, L. *Phys. Rev. Lett.* **85**, 4819–4822 (2000).
3. Bonn, D., Tanaka, H., Wegdam, G., Kellay, H. & Meunier, J. *Europhys. Lett.* **45**, 52–57 (1998).
4. Cipelletti, L., Manley, S., Ball, R. C. & Weitz, D. A. *Phys. Rev. Lett.* **84**, 2275–2278 (2000).

Apoptosis

Baiting death inhibitors

Donald W. Nicholson

Enzymes called caspases that start the process of programmed cell death can be dangerous if activated at the wrong time. A feat of self-restraint keeps one such caspase under control.

Most of the cell death that takes place in mammals occurs by apoptosis — the result of a dedicated biochemical pathway. At the heart of this pathway lies

a family of protein-cleaving enzymes, the caspases, which normally lie dormant in healthy cells and become activated in response to diverse stimuli when cell death

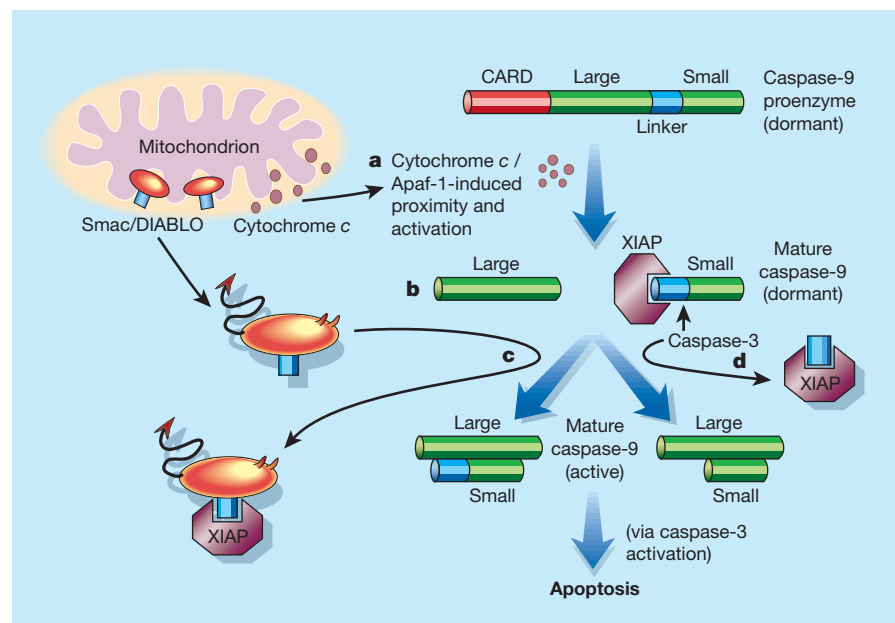


Figure 1 The branch of the programmed-cell-death (apoptosis) pathway that includes caspase-9 and its regulators. a, In response to death stimuli (not shown), a molecule called Apaf-1, together with cytochrome c, which is released from mitochondria, induces the aggregation and processing of caspase-9 proenzymes. The processing includes the removal of the 'CARD' domain, and separation of the large subunit from the linker peptide that connects it to the small subunit. This exposes the linker peptide. b, As shown by Srinivasula *et al.*², one end of the linker peptide mimics the XIAP-binding peptide of Smac/DIABLO, allowing XIAP — an inhibitor of apoptosis — to bind to mature caspase-9. This keeps caspase-9 in a catalytically dormant state. c, d, This suppression of caspase-9 might be counteracted by two mechanisms — competition for XIAP by the apoptosis-promoting protein Smac/DIABLO (c), or caspase-3-mediated removal of the XIAP-binding peptide from caspase-9 (d).

is required. Once activated, the caspases cleave a small subset of the proteins in the cell, and it is the cumulative effect of these cleavage events that accounts for the physical characteristics of apoptosis¹. By definition, then, the caspases are potentially dangerous molecules, and an elaborate system of checks and balances exists to regulate these harbingers of cellular mortality.

On page 112 of this issue², Srinivasula and colleagues reveal another such regulatory mechanism. They show that a key catalyst in the apoptotic pathway, caspase-9, uses a peptide from the end of one of its own subunits as bait to attract a molecule from the IAP (inhibitor-of-apoptosis protein) family. In doing so, caspase-9 remains catalytically dormant even though the initial steps in committing it to its active state have taken place. This mechanism may work to eliminate the untoward effects that small amounts of activated caspase-9 would have on a cell that is not fully committed to apoptotic death.

Dormant caspases exist as precursor polypeptides, or 'proenzymes', which are activated largely by proteolytic processing. This involves cleavage of the polypeptides at specific points to separate out the large and small subunits that together make up the active 'mature' enzyme. The proenzymes themselves have low levels of activity, but processing to their mature forms is required to unleash their full catalytic fury. However, because the proenzymes have this low-level activity, they can process each other if brought close enough together. This is the principal mechanism that launches caspase activation through both of the two known major apoptotic pathways.

The 'extrinsic' pathway starts with the binding of certain external death-inducing molecules ('death ligands') to receptors on the cell surface; this results in the aggregation and proximity-induced maturation of caspase-8. The 'intrinsic' cell-death pathway, which responds to most death stimuli, uses an adapter molecule called Apaf-1 to sequester caspase-9 proenzymes and initiate their activation³ (Fig. 1). Apaf-1 itself requires cytochrome *c*, a protein released from mitochondria, to do this. Caspase-8 and caspase-9 are known as 'initiator' caspases. Once activated, both enzymes feed into a common machinery, which is dominated by caspase-3 in most types of cell.

Once launched, however, the caspase-9/caspase-3 apoptotic pathway does not necessarily proceed unopposed: it is regulated further by proteins of the IAP family. These proteins contain a variable number of so-called BIR domains, regions that directly inhibit caspases. For example, XIAP, the prototypical IAP protein, blocks apoptosis through its BIR3 and BIR2 domains, which inhibit mature caspase-9 and caspase-3, respectively^{4,5}. So the pres-

ence of XIAP slows down and often (although not always) blocks the apoptotic pathway.

The caspase-inhibiting effects of XIAP are antagonized by another apoptosis-promoting mitochondrial protein, called Smac/DIABLO^{6,7}. The amino terminus of Smac/DIABLO binds to a surface groove in the BIR3 domain of XIAP, preventing the inhibition of caspases and of apoptosis⁸⁻¹⁰. In this 'paper wraps stone blunts scissors' approach to regulating cell death¹¹, the ability of caspase-9 and caspase-3 to commit a cell to apoptotic death appears to depend on the dynamics and stoichiometry of XIAP and its antagonist, Smac/DIABLO. Other regulatory polypeptides, such as XIAP-associated factor-1 (ref. 12), may also contribute to these dynamics.

There is no doubt some leakage in this finely tuned regulatory system. But the consequences of even low levels of an unregulated initiator caspase could be devastating to a cell. So it appears that caspase-9 has a back-up strategy: it actually recruits its own inhibitor. Srinivasula *et al.*² show that the amino terminus of the small subunit of mature caspase-9 contains a four-amino-acid peptide that is remarkably similar to the key amino terminus of mammalian Smac/DIABLO and the apoptosis-promoting fruitfly proteins Reaper, Hid and Grim. However, unlike the peptide in Smac/DIABLO, the peptide in mature caspase-9 does not block the caspase-inhibiting effects of XIAP. Instead, this peptide helps XIAP to bind and inhibit caspase-9. The relatively dormant caspase-9 proenzyme (in which this peptide is masked) does not bind XIAP. But once caspase-9 takes on its mature form and the amino terminus of its small subunit is liberated, it uses the exposed peptide as bait to recruit XIAP.

But does this make sense in a cell that may need to fully activate its 'intrinsic' apoptosis pathway? As it turns out, all else holds true. Smac/DIABLO competes with the small subunit of mature caspase-9 for XIAP, and can tear XIAP away. The following scenario may thus play out (Fig. 1). During the normal stresses of cellular existence within a multicellular organism, small amounts of caspase-9 may become activated by self-catalytic or other mechanisms. This rogue protease is inherently dangerous, so the newly exposed amino terminus of its small subunit binds to the BIR3 domain of XIAP; the BIR3 domain in turn inhibits caspase-9. However, under circumstances in which cell death should take place, signals emanating from mitochondria help to reverse or overcome this process. In other words, Smac/DIABLO competes for binding to XIAP and relieves the inhibition of mature caspase-9, while cytochrome *c* stimulates Apaf-1-mediated activation of the broader pool of caspase-9 proenzymes.

Daedalus

David Jones

We are delighted to announce that David Jones, author of the Daedalus column, is back in action. The column returns next week.

Although it was not demonstrated by Srinivasula *et al.*, the XIAP bound to the small subunit of caspase-9 may also be liberated by further proteolytic clipping of the small subunit. We know that caspase-3 cleaves this subunit at a position downstream of the XIAP-binding peptide, releasing the peptide and, presumably, the bound XIAP too. In support of the idea that XIAP remains bound to this peptide even when it is cleaved from the small subunit, Srinivasula *et al.* showed that a synthetic form of the peptide antagonizes the interaction between XIAP and caspase-9, in much the same way that equivalent Smac/DIABLO peptides can block the interaction between XIAP and Smac/DIABLO^{8,9}.

From a therapeutic standpoint, the broader implications of these observations are exciting. If synthetic molecules could be made to mimic the XIAP-binding peptides found in the amino termini of caspase-9, Smac/DIABLO, Reaper, Hid and Grim, those molecules could be used to attenuate IAP-mediated inhibition of caspases. The synthetic molecules might not necessarily provoke apoptosis, but they might sensitize cells to apoptotic stimuli. Unimpeded by IAP proteins, caspases would be more readily activated when appropriately provoked. Cancerous cells often have higher-than-normal levels of caspase proenzymes, and this fact, in combination with a lack of inhibition by IAP proteins, might open a window of opportunity for treating cancer sufferers.

Donald W. Nicholson is at the Merck Frosst Centre for Therapeutic Research, PO Box 1005, Pointe Claire-Dorval, Quebec H9R 4P8, Canada.

e-mail: donald_nicholson@merck.com

- Nicholson, D. W. *Cell Death Differ.* **6**, 1028-1042 (1999).
- Srinivasula, S. M. *et al. Nature* **410**, 112-116 (2001).
- Roy, S. & Nicholson, D. W. *J. Exp. Med.* **192**, 21-26 (2000).
- Deveraux, Q. L. *et al. EMBO J.* **18**, 5242-5251 (1999).
- Ekert, P. G., Silke, J., Hawkins, C. J., Verhagen, A. M. & Vaux, D. L. *J. Cell Biol.* **152**, 483-490 (2001).
- Du, C., Fang, M., Li, Y., Li, L. & Wang, X. *Cell* **102**, 33-42 (2000).
- Verhagen, A. M. *et al. Cell* **102**, 43-53 (2000).
- Chai, J. *et al. Nature* **406**, 855-862 (2000).
- Liu, Z. *et al. Nature* **408**, 1004-1008 (2000).
- Wu, G. *et al. Nature* **408**, 1008-1012 (2000).
- Green, D. R. *Cell* **102**, 1-4 (2000).
- Liston, P. *et al. Nature Cell Biol.* **3**, 128-133 (2001).

Frustrations of fur-farmed mink

Mink may thrive in captivity but they miss having water to romp about in.

Animals may suffer in captivity if they are strongly motivated to perform activities that their housing does not permit. Here we investigate to what extent these limitations affect caged mink and find that these animals will not only pay high costs to be able to perform a range of natural behaviours, but they will also release the 'stress' hormone cortisol when prevented from indulging in swimming, their favourite activity. Despite arguments that mink housed in fur farms have successfully adapted to captivity, these animals may suffer by being deprived of resources that exist in the wild.

Fur farming is widespread in North America, Scandinavia and Europe, with some 30 million mink pelts being produced annually worldwide. On these farms, American mink (*Mustela vison*) are kept in wire-mesh cages (dimensions are typically $0.9 \times 0.4 \times 0.3$ m), with access provided to a single nest box, drinking water and paste-like food. It has been claimed that this causes frustration to the animals¹, which in the wild would patrol territories 1–4 km long, use several nest sites, and hunt by following scent trails, investigating burrows, and diving and swimming for aquatic prey². Others argue that the excellent health and breeding success of farmed mink are evidence of successful adaptation to captivity^{3,4}.

We have objectively investigated this issue of possible deprivation by measuring the costs paid by farm-raised mink to reach resources that will enable them to behave naturally. Because of the key role of pleasure in motivating preference^{5,6}, the results should pinpoint the activities that are important for the welfare of farmed mink.

Eight male and eight female mink were individually housed in closed-economy⁷ set-ups, each consisting of a conventional farm cage, plus seven similarly sized resource compartments. These compartments contained, respectively: a water pool measuring about $1.5 \text{ m} \times 0.5 \text{ m}$ and filled with 0.2 m water (Fig. 1); a raised platform, reached by a 2-m vertical wire tunnel; novel objects such as traffic cones and packaging, which were changed daily; an alternative nest site (a box of hay); toys for manipulation and chewing (tennis balls, for example); and a plastic tunnel. The seventh compartment was left empty to control for the importance of simply making extra space available. Costs to 'pay' to reach the new facilities were imposed by weighting one-way entrance doors by 0, 0.25, 0.5, 0.75, 1 or 1.25 kg for seven successive days.

The animals' activity in each compartment was automatically recorded throughout the day and night, allowing us to calculate four measures of value (Table 1). We found that the animals rated the water pool as the most valuable resource: it attracted the greatest total expenditure and had the highest reservation price, greatest consumer surplus measures of utility, and the most inelastic demand (Table 1).

We next tested the reactions of seven male and seven female mink to having their access blocked to four resources, for 24 hours each, by monitoring their levels of cortisol, which is produced in response to stress⁸. The resources chosen were of high, intermediate or low value, as judged by the cost the animals would pay to reach them (the water pool, alternative nest site or empty compartment, respectively), and the fourth was an



Figure 1 A caged American mink (*Mustela vison*). This one has access to a water pool, a highly valued resource enabling the animal to indulge its natural instinct for diving and swimming.

essential physiological resource — food. During each trial, urine was collected for assay of excreted cortisol; levels of creatinine were also determined to correct for any differences in urine concentration.

When deprived of food, the minks' urinary cortisol increased by $50.0 \pm 16.1\%$ over baseline (paired $t = 2.77$, d.f. = 13, $P < 0.05$); it increased by $33.8 \pm 11.2\%$ when access to the water pool was denied ($t = 2.75$, $P < 0.05$). These two effects did not differ significantly ($t = 2.47$, $P > 0.05$). Cortisol excretion was not increased by blocking the other two resources. Consistent with cortisol's metabolic functions⁸, the animals' excretion of the hormone in response to being denied access to food or to the water pool correlated with an increase in their physical activity (food: $R = 0.49$, $P = 0.07$; water pool: $R = 0.70$, $P = 0.005$; $n = 14$). Overall, however, changes in activity only reached significance during food deprivation ($t = 2.45$, d.f. = 13, $P < 0.05$).

Table 1 Value to farmed mink of resources allowing natural behaviour

Resource	Elasticity of demand	Consumer surplus (kg)		Reservation price (kg)	Total expenditure (kg)
		Travel cost	Aggregate		
Water pool	$0.26 \pm 0.04^{\dagger}\P$	$81.41 \pm 9.97^{\dagger}\P$	24.00	$1.25 \pm 0.00^{\dagger}\P$	$134.33 \pm 11.18^{\dagger}\P$
Alternative nest site	$0.41 \pm 0.08^{\dagger}\P$	$60.72 \pm 5.67^{\dagger}\P$	22.75	$1.17 \pm 0.05^{\dagger}\P$	$114.83 \pm 13.27^{\dagger}\P$
Novel objects	$0.58 \pm 0.08^{\dagger}\P$	$54.58 \pm 5.02^{\dagger}\P$	22.50	$1.16 \pm 0.04^{\dagger}\P$	$83.62 \pm 9.93^{\dagger}\P$
Raised platform	$0.57 \pm 0.07^{\dagger}\P$	$50.87 \pm 7.65^{\dagger}\P$	22.25	$1.14 \pm 0.06^{\dagger}\P$	$82.17 \pm 16.11^{\dagger}\P$
Toys	$0.62 \pm 0.05^{\dagger}\P$	$24.30 \pm 3.25^{\dagger}\P$	21.00	$1.06 \pm 0.07^{\dagger}\P$	$34.11 \pm 6.39^{\dagger}\P$
Tunnel	$0.73 \pm 0.07^{\dagger}\P$	$21.61 \pm 1.72^{\dagger}\P$	20.75	$1.06 \pm 0.06^{\dagger}\P$	$26.33 \pm 3.66^{\dagger}\P$
Empty cage	$0.77 \pm 0.06^{\dagger}\P$	$9.19 \pm 0.90^{\dagger}\P$	17.00	$0.84 \pm 0.07^{\dagger}\P$	$8.76 \pm 1.34^{\dagger}\P$

The 'price elasticity of demand', widely advocated for assessing animal welfare⁹, was calculated as the gradient of the log-log plot of visit price versus visit number for each resource⁹. 'Consumer surplus', used by human welfare economists to assess resource value¹⁰, was calculated by measuring the area under two types of demand curve: a plot of visit price versus visit number, analogous to the 'travel cost method' of environmental economists¹¹, and an aggregate plot of price versus the number of subjects prepared to pay each price⁹. 'Reservation price'⁹, akin to the 'break point'¹² of experimental psychologists, was calculated as the maximum price paid to reach each resource. 'Total expenditure' per unit time, as given precedence by behavioural ecologists¹³, was also calculated, the values shown here being those calculated for the six weeks of the experiment. All values are given as means and standard errors ($n = 16$), except for 'aggregate consumer surplus'. Low values for elasticity of demand indicate little decline in visit rate as visit costs increase; for all remaining measures, a high numeric value indicates a high usage value.

[†]†§ Resources whose values differ significantly at $P < 0.01$ (Tukey's t -test).

[†] Resources that significantly contribute (at $P < 0.01$) to a general linear model (Minitab 12) of the form 'value equals mink (sex) + sex + resource + sex $\times$ resource', where 'mink' is a random factor.

Our results indicate that fur-farmed mink are still motivated to perform the same activities as their wild counterparts, despite being bred in captivity for 70 generations⁴, being raised from birth in farm conditions, and being provided with food *ad libitum*. The high level of stress experienced by mink denied access to the pool, rated as the most valuable resource, is evidenced by an increase in cortisol production indistinguishable from that caused by food deprivation. These results suggest that caging mink on fur farms does cause the animals frustration, mainly because they are prevented from swimming.

Georgia J. Mason*, Jonathan Cooper†, Catherine Clarebrough*

*Zoology Department, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
e-mail: georgia.mason@zoology.oxford.ac.uk

Structural colour

Now you see it — now you don't

The dazzling iridescence seen in some hummingbirds¹ and tropical butterflies² arises from natural optical phenomena, the brightest of which originate in nanoscale structures that produce ultra-high reflectivity and narrow-band spectral purity³. Here we investigate the coloration of male *Ancyluris meliboeus* Fabricius butterflies⁴, which have patches of unusual microstructure on their ventral wing scales. We find that this highly tilted, multilayered arrangement produces a bright iridescence of broad wavelength range and generates a strong flicker contrast from minimal wing movement.

The microstructure in the iridescent scales of *A. meliboeus* comprises multilayers of cuticle and air within a discrete *Morpho*-like² ridging that runs the length of each scale (Fig. 1a). These layers are tilted at about 30 degrees to the base of the scale, causing two distinct optical effects. In the first, abrupt termination of each cuticle layer at the upper ridge surface presents a strong periodicity of about 700 nm, which contributes a diffractive component.

We analysed the structure's reciprocal space⁵ to find out how the periodicity of the multilayers and diffracting elements scatter incident light. The diffractive component appears to combine additively with the interference from the underlying multilayer to produce a broad range of coloration, as well as a limited reverse colour change with angle compared to that associated with conventional flat multilayering.

The second, more striking effect arising from the tilted multilayering accounts for the strongly bistable nature of the wing

†School of Agriculture and Horticulture, De Montfort University, Caythorpe Court, Caythorpe, Lincolnshire NG32 3EP, UK

1. Animal Aid *The Fur Trade Factsheet* (Animal Aid, Tonbridge, 1999).
2. Dunstone, N. *The Mink* (Poyser Natural History, London, 1993).
3. The International Fur Trade Federation *Fur Farming* (<http://www.iftf.com/farming/html>).
4. Joergenson, G. (ed.) *Mink Production* (Scientific, Hillerød, 1985).
5. Dawkins, M. *Behav. Brain Sci.* **13**, 1–61 (1990).
6. Rolls, E. *Brain. Behav. Sci.* **23**, 177–200 (2000).
7. Hursh, S. R. *J. Exp. Anal. Behav.* **34**, 219–238 (1980).
8. Toates, F. *Stress — Conceptual and Biological Aspects* (Wiley, Chichester, 1995).
9. Varian, H. *Intermediate Microeconomics — A Modern Approach* (Norton, New York, 1996).
10. Ng, Y.-K. *Welfare Economics* (Macmillan, Basingstoke, 1983).
11. Hartwick, J. M. & Olewiler, N. D. *The Economics of Natural Resource Use* (Addison-Wesley, Reading, MA, 1998).
12. Reilly, S. *Pharmacol. Biochem. Behav.* **63**, 301–311 (1999).
13. Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton Univ. Press, NJ, 1986).

reflectivity in diffuse white light: it is either 'on', when an observer sees one of a broad range of colours, or it is 'off' and

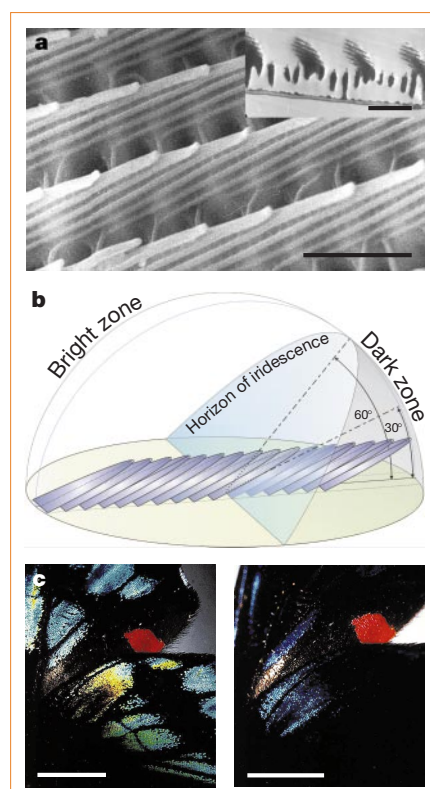


Figure 1 Tilted multilayer ridging divides the observation hemisphere above the iridescent scales of the butterfly *A. meliboeus* to produce strong colour flicker. **a**, Scanning electron microscope image of the surface of an iridescent scale. Inset, transmission electron microscope image of the cross-section through an iridescent scale at 45° to the line of ridging. Scale bar, 1 μm (inset, 2 μm). **b**, Tilt-induced bright and dark zones in the observation hemisphere over an iridescent scale. **c**, Real-colour images showing the same portion of the butterfly's wing under diffuse illumination: left, image with the camera in the wing's bright zone; right, image after moving the camera 15° round the observation hemisphere into the dark zone (the red region is coloured by pigmentation). Scale bars, 4 mm.

produces no reflected iridescence. The 30-degree layer tilt causes a 60-degree portion of the wing's 'observation hemisphere' (Fig. 1b) not to appear iridescent ('dark zone' in Fig. 1b).

Over the remaining 120 degrees of the hemisphere, diffuse light produces iridescent reflection. Under identical illumination conditions, other structurally coloured insects and animals are seen as iridescent at angles over the entire hemisphere above their reflecting surfaces.

This structural arrangement is important in signalling by the butterfly. On or near the edge of the *A. meliboeus* dark zone, wing movements of no more than a few degrees generate ultra-high-contrast colour flicker in reflectivity (Fig. 1c). In species whose observation hemispheres have no dark zone, wing movements of large amplitude are necessary to achieve colour flicker.

Intermittent colour flicker is a useful conspecific trigger stimulus⁶, which becomes increasingly supernormal at higher and higher frequencies until the flicker-fusion frequency of the observer's visual system is reached⁷. However, inertial and physiological constraints usually prevent the wings from producing flicker signals at frequencies approaching those of visual flicker fusion⁶.

Ancyluris meliboeus avoids the need for large wing movements by means of this ingenious optical structure. By virtue of strong diffracting elements above the multilayering on its iridescent scales, this butterfly produces a broader colour range than would be accessible through interference alone — hence its alias as a "living jewel"⁸.

P. Vukusic*, J. R. Sambles*, C. R. Lawrence†, R. J. Wootton‡

*Thin Film Photonics, School of Physics, Exeter University, Exeter EX4 4QL, UK
e-mail: p.vukusic@ex.ac.uk

†Mechanical Sciences Sector, DERA, Farnborough GU14 0LX, UK

‡School of Biological Sciences, Exeter University, Exeter EX4 4PS, UK

1. Greenwalt, C. H., Brandt, W. & Friel, D. D. *J. Opt. Soc. Am.* **50**, 1005–1013 (1960).
2. Vukusic, P., Sambles, J. R., Lawrence, C. R. & Wootton, R. J. *Proc. R. Soc. Lond. B* **266**, 1402–1411 (1999).
3. Land, M. F. *Progr. Biophys. Molec. Biol.* **24**, 75–106 (1972).
4. DeVries, P. J. *The Butterflies of Costa Rica and Their Natural History II: Riodinidae* (Princeton Univ. Press, NJ, 1997).
5. Guinier, A. *X-Ray Diffraction* (Freeman, San Francisco, 1963).
6. Magnus, D. B. E. *Proc. Int. Congr. Entomol.* **2**, 405–418 (1958).
7. Silberglied, R. E. in *The Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. E.) 207–223 (Symp. R. Soc. 11, Academic, London, 1984).
8. Smart, P. *Encyclopedia of the Butterfly* (Transworld, London, 1976).

erratum

A viable herd of genetically uniform cattle

P. M. Visscher *et al.*

Nature **409**, 303 (2001)

The correct name of the last author is J. L. Williams.

Mammalian MAP kinase signalling cascades

Lufen Chang & Michael Karin

Department of Pharmacology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636, USA

Mitogen-activated protein kinases (MAPKs) are important signal transducing enzymes, unique to eukaryotes, that are involved in many facets of cellular regulation. Initial research concentrated on defining the components and organization of MAPK signalling cascades, but recent studies have begun to shed light on the physiological functions of these cascades in the control of gene expression, cell proliferation and programmed cell death.

MAPKs are evolutionary conserved enzymes connecting cell-surface receptors to critical regulatory targets within cells. MAPKs also respond to chemical and physical stresses, thereby controlling cell survival and adaptation. MAPK activity is regulated through three-tiered cascades composed of a MAPK, MAPK kinase (MAPKK, MKK or MEK) and a MAPKK kinase or MEK kinase (MAPKKK or MEKK)¹. These modules may be activated by STE20 kinases or small GTP-binding proteins². Many MAPKs activate specific effector kinases—MAPK-activated protein kinases (MAPKAPKs)—and are inactivated by MAPK phosphatases.

As the regulation of MAPK signalling has been discussed elsewhere^{1,2}, we focus on recent progress in understanding MAPK function in mammals. Because of the ease of genetic analysis, the functions and regulation of MAPK cascades are best understood in budding yeast³, but the analysis of targeted mutations in mice and development of specific inhibitors have begun to shed light on their functions in mammals. It is becoming clear that MAPKs regulate almost all cellular processes, from gene expression to cell death.

Mammalian MAPKs

Mammals express at least four distinctly regulated groups of MAPKs, extracellular signal-related kinases (ERK)-1/2, Jun amino-terminal kinases (JNK1/2/3), p38 proteins (p38 α /p38 β /p38 γ /p38 δ) and ERK5, that are activated by specific MAPKKs: MEK1/2 for ERK1/2, MKK3/6 for the p38, MKK4/7 (JNKK1/2) for the JNKs, and MEK5 for ERK5. Each MAPKK, however, can be activated by more than one MAPKKK, increasing the complexity and diversity of MAPK signalling. Presumably each MAPKKK confers responsiveness to distinct stimuli. For example, activation of ERK1/2 by growth factors depends on the MAPKKK c-Raf, but other MAPKKKs may activate ERK1/2 in response to pro-inflammatory stimuli.

The situation is quite complex with the JNK and p38 cascades, which respond to many stimuli and can be activated on overexpression of at least a dozen MAPKKKs, whose physiological functions and specificities remain elusive. Dominant-negative catalytically inactive MAPKKKs can be used to address these issues, but the results of such experiments can be confusing. For example, at least four MAPKKKs were implicated in JNK activation by tumour necrosis factor (TNF) and interleukin (IL)-1, including ASK1 (MEKK5), TAK1, MLK and MEKK1 (reviewed in ref. 4). These difficulties are probably generated by loss of specificity caused by overexpression of dominant-negative mutants.

Systematic genetic ablation of MAPKKKs may provide definitive answers to many of these questions. For instance, it has been shown that MEKK1 is required for JNK activation in embryonic stem cells or fibroblasts by microtubule-disrupting agents, lysophosphatidic acid, viral infection, double-stranded RNA and serum^{4,5}. MEKK1 also is required for activation of JNK by TNF and IL-1 in embryonic stem cells⁴, but not in fibroblasts⁵. These experiments revealed that MEKK1 is dispensable for activation of JNK by ultraviolet radiation and protein-synthesis inhibitors and is not essential for ERK1/2 or

p38 activation. Such results suggest considerable specificity in MAPK activation even at the MAPKKK level.

Specificity in MAPK activation and function

The control of diverse cellular processes in response to a plethora of extracellular stimuli by a few MAPKs implies that considerable specificity is built into MAPK activation and function. Several distinct but not mutually exclusive mechanisms secure specificity in MAPK activation. In yeast, the STE5 protein functions as a scaffold that organizes the three components of a pheromone-responsive MAPK cascade and its upstream activators into a specific module³. A search for analogous mammalian scaffolds led to JIP1, which organizes JNK1/2, MKK7 and the MAPKKK MLK1 into a specific signalling cassette (Fig. 1a)⁶. MP1, another mammalian scaffold protein not related to JIP-1, interacts with ERK1 and MEK1, thereby potentiating ERK1 activation⁷.

A second mechanism ensuring specific MAPK activation depends on sequential physical interactions between members of a given cascade⁸. For instance, JNK1/2 is bound by the N-terminal extension of MKK4 (JNKK1), which also interacts with the catalytic domain of MEKK1. Each interaction is disrupted on activation of the downstream kinase (Fig. 1b). Whereas scaffolds increase specificity at the expense of signal amplification, sequential interactions provide amplification without sacrificing specificity⁸. Both mechanisms may operate in parallel, leading to differential activation of the same MAPK in response to distinct stimuli.

A third and more complex mechanism for signal amplification is based on the ability of MAPKs to regulate indirectly the expression of both ligands and inhibitors for cell-surface receptors that feed into MAPK cascades⁹. Such positive and negative autocrine loops can generate specific spatial patterns of MAPK activation (Fig. 1c).

Once activated, MAPKs need to find their targets. Although it is necessary to limit the phosphorylation of irrelevant substrates, it is also important that each MAPK recognizes a number of substrates to allow the regulation of many processes. This problem seems to have been solved through a bipartite enzyme–substrate interaction. All MAPKs recognize similar phosphoacceptor sites composed of serine or threonine followed by a proline, and the amino acids that surround these sites further increase the specificity of recognition by the catalytic pocket of the enzyme. Full specificity is ensured through a docking interaction mediated by another site on the kinase that recognizes a distinct site on the substrate (Fig. 1d). As first described for c-Jun, a short sequence that precedes the two principal JNK phosphoacceptor sites, Ser 63 and Ser 73, is recognized by an interaction surface on JNK, outside its catalytic pocket¹⁰. This interaction determines the efficiency of c-Jun phosphorylation and the choice of phosphoacceptor sites (Fig. 1d). Similar interactions involving distinct phosphoacceptor and docking sites determine substrate recognition by other MAPKs¹¹.

Spatial localization of signalling molecules further augments specificity in signal transduction. For instance, MEKK1 colocalizes with elements of the cytoskeleton¹, and cytoskeletal rearrangements

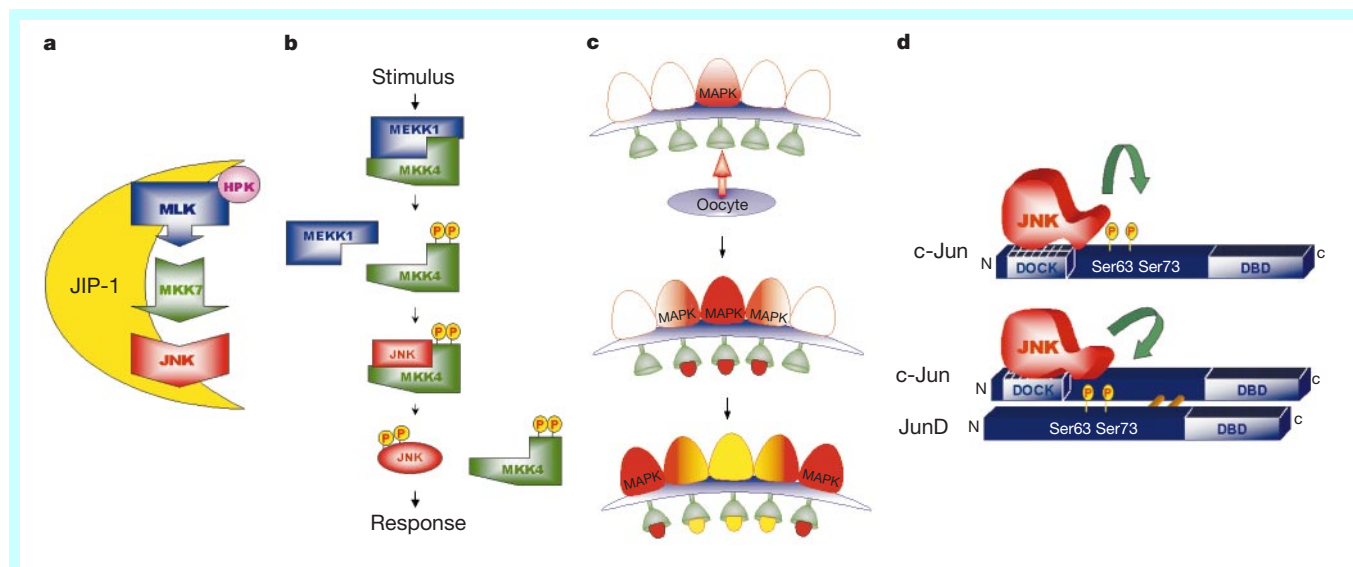


Figure 1 Mechanisms for generation of specificity in MAPK activation and function. **a**, JIP-1 acts as a scaffold that ties together the upstream kinase HPK-1, the MAPKKK MLK1, the MAPKK MKK7 and the MAPK JNK into a specific signalling module. **b**, Sequential and specific interaction between members of a MAPK cascade. MEKK1 interacts with inactive MKK4 to form a MEKK1–MKK4 complex. MEKK1 phosphorylates and activates MKK4, resulting in dissociation of the complex. The free and active MKK4 then specifically interacts with JNK. Once JNK is activated, the MKK4–JNK complex is disrupted and the active JNK is now freed to phosphorylate downstream effectors. **c**, Specific spatial patterns of MAPK activation by positive and negative autocrine loops in the *Drosophila* oocyte. A ligand called Gurken (red arrow) that is localized in the underlying oocyte activates a MAPK cascade (red gradient) in follicle cells, through an EGF receptor (green). This results in upregulation of another EGF receptor agonist, 'Spitz' (red), that activates MAPK in neighbouring cells. This is followed by expression of the inhibitory ligand 'Argos' (yellow) in midline cells with higher levels of MAPK activity, which eventually inhibits of MAPK activity in this region. **d**, The JNK docking region (DOCK) of c-Jun determines its specific phosphorylation at Ser 63 and Ser 73 by JNK. A substrate lacking a functional docking site, such as JunD, can be phosphorylated by JNK that is recruited through the DNA-binding domain (DBD) of c-Jun.

may stimulate MEKK1 activity^{4,5}. Such localized JNK activation may allow MEKK1 to modulate cell motility^{4,5}.

General functions of MAPK cascades

One of the most explored functions of MAPK signalling modules is regulation of gene expression in response to extracellular stimuli¹². JNKs phosphorylate Jun proteins and thereby enhance their ability to activate transcription without affecting DNA binding¹⁰. Most MAPKs phosphorylate Ets transcription factors that are involved in induction of *fos* genes, whose products heterodimerize with Jun proteins to form activation protein 1 (AP-1) complexes¹². The p38 proteins phosphorylate and enhance the activity of MEF2C and related family members¹³. In all of these cases, the MAPKs function inside the nucleus and target transcription factors that are pre-bound to DNA. Most transcription factors that are regulated by MAPKs are dimers, and structural analysis indicates that some MAPKs dimerize on activation¹; however, MAPK dimerization may facilitate phosphorylation of dimeric transcription factors.

Although transcription factors are important MAPK targets, only part of the active MAPK pool translocates to the nucleus and much remains in the cytoplasm and other subcellular compartments. Correspondingly, MAPKs also regulate gene expression through post-transcriptional mechanisms involving cytoplasmic targets. JNK, for instance, is involved in IL-2 messenger RNA stabilization in activated T cells¹⁴. The target for JNK action appears to be a cytoplasmic protein that is recruited to IL-2 mRNA through two other proteins that bind to the 5' untranslated region¹⁴. Other mRNA species are stabilized in response to p38 activation^{15,16}. MAPKs are also involved in translational control, but in this case (as well as in some cases of mRNA stabilization¹⁵) their action is mediated by the MAPKAPK-2 and MNK1 effector kinases^{17,18}. MNK1 phosphorylates translation initiation factor 4e (eIF-4e), augmenting the interaction of the cap-binding complex with capped mRNAs, thereby enhancing polysome assembly.

MAPKs regulate cell proliferation. ERK1/2 stimulates DNA synthesis through phosphorylation of carbamoyl phosphate synthetase

II, a rate-limiting enzyme in pyrimidine nucleotide biosynthesis¹⁹. In addition, the ERKs, probably through MAPKAPKs such as RSK, promote cell-cycle progression by inactivating MYT1, a cell-cycle inhibitory kinase²⁰, but arrest meiotic cells at metaphase II by activating a cytostatic factor^{21,22}. The ERKs can also stimulate cell proliferation indirectly by enhancing AP-1 activity, resulting in cyclin D1 induction²³. That pathway alone is insufficient for initiation of DNA synthesis and needs assistance from phosphatidylinositol-3-OH kinase, which is indirectly activated by autocrine growth factors whose expression is ERK responsive²³. Autocrine factors are important in cellular responses to MAPKs⁹ and can provide a means for one MAPK cascade to activate other signalling pathways²⁴. In T cells, the JNK pathway may stimulate proliferation by inducing IL-2 expression²⁵.

MAPKs also control cell survival. By and large, activation of ERK1/2 has been linked to cell survival, whereas JNK and p38 are linked to induction of apoptosis²⁶. This dichotomy, however, is an oversimplification, and the actual roles of each MAPK cascade are highly cell type and context dependent. A pro-apoptotic function has been established for prolonged JNK activation in certain neuronal cells, especially in the CA1 layer of the hippocampus after intense activation of glutamate receptors^{26–28}. This process does not occur in *Jnk3*^{−/−} mice²⁸ or in mice expressing a mutant form of c-Jun (c-Jun^{A63/73}) lacking the JNK phosphoacceptor sites²⁹; however, transient and more modest JNK activation in response to physiological stimulation does not result in neuronal apoptosis³⁰. JNK-induced neuronal apoptosis partially depends on induction of death cytokines, such as Fas ligand, whose promoter contains an AP-1 site, explaining the dependence on c-Jun phosphorylation^{27,29}. JNK1 and JNK2 are required for apoptosis of ultraviolet-irradiated fibroblasts, in which they have been suggested to activate directly the apoptotic machinery³¹. However, the combined JNK1/2 deficiency only prevents ultraviolet-induced apoptosis and does not hinder other pro-apoptotic stimuli³¹. c-Jun-deficient fibroblasts also exhibit specific resistance to ultraviolet-induced apoptosis³². Given that JNKs are vigorously activated after ultraviolet exposure and needed

for c-Jun induction³¹, it is likely that the pro-apoptotic function of JNK1/2 in ultraviolet-irradiated fibroblasts is mediated through c-Jun³². Nevertheless, in cells exposed to physiologically relevant doses of ultraviolet radiation, JNK-induced c-Jun stimulates cell-cycle re-entry rather than inducing apoptosis³².

The mechanism by which ERK protects cells from apoptosis is complex and Ras, a potent ERK activator, may also promote apoptosis³³. In cerebellar granular cells, ERK activation by survival factors prevents apoptosis through RSK, which inactivates the pro-apoptotic protein BAD³⁴. ERK may also induce growth factors that promote cell survival.

In vivo analysis of MAPK and MAPKK function

Although much can be learned about MAPK functions from *ex vivo* studies, it is important to investigate both their normal and their pathophysiological functions in the whole animal. Such analysis, accomplished through generation of mutant mouse strains by gene targeting, is also useful for determining the organization and regulation of MAPK signalling cascades. As the different MAPKKs exhibit high levels of specificity, it should also be possible to learn about MAPK function by targeting relevant MAPKKs. Some of the findings and the conclusions derived from analyses of MAPK- and MAPKK-deficient mice are summarized in Table 1.

ERK1/2. Only the knockout of ERK1 has been described³⁵. *Erk1*^{-/-} mice are viable and appear normal and with a modest defect in T-cell development (positive selection)³⁵. It is likely that most ERK1 functions are equally served by ERK2. A similar and more marked defect is present in transgenic mice expressing dominant-negative MEK1 in thymocytes³⁶; and *Mek1*^{-/-} mice die *in utero*, exhibiting defective placental vascularization³⁷.

Despite impairment of positive selection and loss of ERK activity, thymocytes expressing dominant-negative MEK1 exhibit normal IL-2 induction and a mitogenic response³⁶. Thus, other MAPKs are probably involved in IL-2 induction. As suggested by the *Mek1*^{-/-} phenotype, MEK1 is required for migration of vascular endothelial cells as well as fibroblasts plated on fibronectin³⁷. However, fibronectin-induced ERK activity remains normal³⁷. An ERK1/2-MEK1 module may target specific substrates involved in cell migration while making only a small contribution to total ERK1/2 activity.

JNKs. All three *Jnk* loci have been knocked out. On mixed genetic backgrounds, none of the mutations results in lethality or obvious defects. However, *Jnk1*^{-/-}*Jnk2*^{-/-} double mutants die at mid-gestation, exhibiting defective neural-tube closure^{38,39}. Thus, JNK functions needed for development and viability are not isoform specific. Unexpectedly, deletion of MKK4 results in a more severe phenotype than the combined loss of JNK1/JNK2: mid-gestational lethality caused by abnormal liver development⁴⁰. The same phenotype is caused by complete loss of c-Jun⁴².

Although JNK1 and JNK2 are widely expressed, much attention has been paid to their functions in T cells. Unfortunately, owing to different experimental protocols and genetic backgrounds, the results of these MKK4 knockout experiments are confusing. In T thymocytes, JNK activity is synergistically stimulated by occupancy of the T-cell receptor and the CD28 auxiliary receptor and this activation parallels IL-2 induction⁴². Induction of anergy, which prevents T-cell activation, interferes with JNK activation⁴³. Congruently, two groups found defective IL-2 production and proliferation after co-stimulation in JNK1/2- or MKK4-deficient T cells^{44,45}.

This defect was also observed in *Jnk1*^{+/-}*Jnk2*^{+/-} double heterozygote T cells⁴⁴, suggesting that partially reduced JNK1/2 activity—which is expected to occur in *Mkk4*^{-/-} cells—is sufficient to decrease T-cell activation. At any rate, this defect is partial and bypassed by strong stimulation or antigen-presenting cells⁴⁴; therefore, it is possible that neither JNK1 nor JNK2 has a major role in controlling T-cell proliferation *in vivo*. Unexpectedly, JNK1/2-deficient T cells exhibit normal AP-1 activity, but display a partial defect in activation of NF-AT, another transcription factor required for IL-2 induction⁴⁵.

Table 1 Phenotypes of MAPKK and MAPK knockout mice

MAPKK/MAPK	Phenotypes	Similar to
MEK1	Defective placental vascularization ³⁷	ERK2?
MKK4 MKK7	Defective liver development ⁴⁰ Embryonic lethality of unknown cause ⁴⁸	c-Jun knockout ⁴¹
MKK3	Defective IL-12 production ⁵²	
ERK1	Defective T-cell development (positive selection) ³⁵	MEK1 dn negative transgenics
JNK1	Defective T-cell differentiation to Th2 cells ⁴⁶	
JNK2	Defective T-cell differentiation to Th1 cells ⁴⁷	
JNK1 or JNK2	Defective T-cell proliferation and IL-2 production ²⁵	JNK1 dn negative transgenics MKK4 knockouts
JNK1 or JNK2	Defective activation induced death of thymocytes ²⁵	JNK1 dn negative transgenics
JNK1 & JNK2	IL-2 overproduction ⁴⁸	MKK7 knockout
JNK1 & JNK2	Neural tube closure ^{38,39}	
JNK3	Resistance to excitotoxic neuronal cell death ²⁸	c-Jun ^{A63/73} knockin ²⁹
p38α	Placental defect ⁵¹ (trophoblast cells)	
p38α	Insufficient production of erythropoietin ⁵⁰	

dn, dominant-negative

Another group, however, found that *Jnk1*^{-/-} or *Jnk2*^{-/-} T cells produce IL-2 normally but undergo aberrant differentiation to effector cells^{46,47}. Furthermore, an experiment in which *Jnk1*^{-/-}*Jnk2*^{-/-} double knockout ES cells were used to reconstitute the lymphoid lineage of immunodeficient mice suggests that JNK1/2 negatively regulated IL-2 production⁴⁸.

These results are supported by analysis of *Mkk7*^{-/-} T cells, which exhibit defective JNK activation⁴⁸, but they contradict experiments in which inhibition of JNK activity correlates with decreased IL-2 expression^{42,43}. Although this discrepancy is not fully explainable, each experiment relied on a different approach and *Mkk7*^{-/-} T cells also exhibit much higher p38 activity than wild-type cells⁴⁸. It is therefore worthwhile to evaluate whether the overproduction of IL-2 is sensitive to p38 inhibitors.

Nevertheless, instead of T-cell activation JNK1 and JNK2 may be more intimately involved in formation of effector T cells^{46,47}. As T cells lacking both JNK1 and JNK2 differentiate along the T-helper cell type-2 (Th2) path⁴⁸, the most important function of JNK1/2 in T cells could be to direct non-differentiated cells to the Th1 fate. JNK1/2 may also be involved in negative selection of T cells^{25,44}. JNK-deficient cells, however, remain sensitive to other inducers of apoptosis, providing another example of the cell type- and stimulus-specific involvement of JNK in apoptosis. Surprisingly, no B-cell defects were detected in any of the JNK-deficient mice^{25,44,46–48}, but a partial proliferation defect was found in *Mkk4*^{-/-} B cells⁴⁹.

In contrast to JNK1/2, the JNK3 isoform is expressed predominantly in the nervous system. *Jnk3*^{-/-} mice are resistant to excitotoxic killing of neurons in the CA1 layer of the hippocampus²⁸. The same phenotype is exhibited by c-Jun^{A63/73} mice²⁹. Thus, in this apoptotic pathway as in ultraviolet-irradiated fibroblasts³², the function of JNK is exerted through its effects on gene expression.

p38. The only p38 isozyme whose *in vivo* function has been examined genetically is p38α. Unlike deletions of individual *Jnk* loci or *Erk1*, inactivation of p38α results in embryonic lethality^{50,51}. It is not clear whether the lack of compensation by other isoforms is indicative of distinct biochemical functions or a marked difference in expression patterns. The severity of the phenotype and the cause of death vary with the genetic background in which the p38α deficiency is examined. In an inbred C57Bl6/J background, p38α deficiency results in lethality before day 11 of gestation, owing to defective placental development⁵¹. However, in a mixed background, some p38α^{-/-} mice survive longer⁵⁰.

Analysis of these mice suggests that after passing the earlier hurdle imposed by development of placental trophoblasts, p38α-deficient

mice face a second hurdle owing to insufficient production of erythropoietin (Epo). Experiments with human hepatoma cells—a surrogate for fetal liver, the main site of fetal Epo production—indicate that p38 α is activated in response to hypoxia and that its activation is required for hypoxia-induced stabilization of *Epo* mRNA⁵⁰. Regardless of the molecular processes in which p38 α participates, these experiments show that a MAPK previously thought to be mostly involved in inflammation and stress does have critical developmental functions.

Consistent with the presence of two MAPKKs that activate p38 MAPKs, *Mkk3*^{−/−} mice are viable without obvious abnormalities⁵². Although *Mkk3*^{−/−} macrophages exhibit reduced endotoxin-induced p38 activity, most cytokines in these cells are normally induced following endotoxin exposure with the exception of IL-12 p35 and p40 subunits⁵². Although this defect occurs at the transcriptional level, the p38 responsive transcription factor is unknown.

Conclusions and prospects

In mammals, MAPK signalling cascades regulate important cellular processes including gene expression, cell proliferation, cell survival and death, and cell motility. With a few exceptions (mostly transcription factors), the MAPK substrates that regulate these processes are not known and their identification is a major challenge. The definition of physiologically relevant MAPK substrates should result in a thorough understanding of MAPK action. Currently, the only process in which a target (c-Jun) for a MAPK (JNK3) has been genetically defined is excitotoxic neuronal apoptosis^{28,29}. As revealed by the involvement of c-Jun, this response is mediated through an effect on gene expression.

Similar analysis can be used to determine how many other biological responses to MAPK activation are mediated through changes in gene expression. As different cells express distinct sets of transcription factors, transcriptional control can easily account for much of the cell-type specificity in MAPK action. Techniques such as DNA microarray analysis should enable the definition of genes that are regulated in response to individual MAPKs, whereas proteomic analyses can be applied to the identification of substrates. Finally, specific inhibitors of MAPK signalling are becoming available. In addition to their utility in identifying MAPK functions, it is likely that they will prove useful in treating conditions, such as inflammation or cancer, caused by deregulation of MAPK cascades. □

- English, J. *et al.* New insights into the control of MAP kinase pathways. *Exp. Cell Res.* **253**, 255–270 (1999).
- Gutkind, S. J. Regulation of mitogen-activated protein kinase signalling networks by G-protein coupled receptors [online] (http://www.stke.org/cgi/content/full/OC_sigrtrans/2000/40/re1) (2000).
- Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187–197 (1995).
- Xia, Y. *et al.* MEK kinase 1 is critically required for c-Jun N-terminal kinase activation by proinflammatory stimuli and growth factor-induced cell migration. *Proc. Natl Acad. Sci. USA* **97**, 5243–5248 (2000).
- Yujiri, T., Sather, S., Fanger, C. R. & Johnson, G. L. Role of MEK1 in cell survival and activation of JNK and ERK pathways defined by targeted gene disruption. *Science* **282**, 1911–1914 (1998).
- Whitmarsh, A. J., Cavanagh, J., Tournier, C., Yasuda, J. & Davis, R. J. A mammalian scaffold complex that selectively mediates MAP kinase activation. *Science* **281**, 1671–1674 (1999).
- Schaeffer, H. J. *et al.* MP1: a MEK binding partner that enhances enzymatic activation of the MAP kinase cascade. *Science* **281**, 1668–1671 (1998).
- Xia, Y., Wu, Z., Su, B., Murray, B. & Karin, M. JNK1 organizes a MAP kinase module through specific and sequential interactions with upstream and downstream components mediated by its amino-terminal extension. *Genes Dev.* **12**, 3369–3381 (1998).
- Wasserman, J. D. & Mathew, F. An autoregulatory cascade of EGF receptor signaling patterns the *Drosophila* egg. *Cell* **95**, 355–364 (1998).
- Kallunki, T., Deng, T., Hibi, M. & Karin, M. c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell* **87**, 929–939 (1996).
- Tanoue, T., Adachi, M., Moriguchi, T. & Nishida, E. A conserved docking motif in MAP kinase common to substrates, activators and regulators. *Nature Cell Biol.* **2**, 110–116 (2000).
- Treisman, R. Regulation of transcription by MAP kinase cascades. *Curr. Opin. Cell Biol.* **8**, 205–215 (1996).
- Han, J., Jiang, Y., Li, Z., Kravchenko, V. V. & Ulevitch, R. J. Activation of the transcription factor MEF2C by the MAP kinase p38 in inflammation. *Nature* **386**, 296–299 (1997).
- Chen, C. Y. *et al.* Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation. *Genes Dev.* **14**, 1236–1248 (2000).
- Winzen, R. *et al.* The p38 MAP kinase pathway signals for cytokine-induced mRNA stabilization via MAP kinase-activated protein kinase 2 and an AU-rich region-targeted mechanism. *EMBO J.* **19**, 6742–6753 (1999).

- Lasa, M. *et al.* Regulation of cyclooxygenase 2 mRNA stability by the mitogen-activated protein kinase p38 signaling cascade. *Mol. Cell Biol.* **20**, 4265–4274 (2000).
- Kotlyarov, A. *et al.* MAPKAP kinase 2 is essential for LPS-induced TNF- α biosynthesis. *Nature Cell Biol.* **1**, 94–97 (1999).
- Pyronnet, S. *et al.* Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* **18**, 270–279 (1999).
- Graves, L. M. *et al.* Regulation of carbamoyl phosphate synthetase by MAP kinase. *Nature* **403**, 328–332 (2000).
- Palmer, A., Gavin, A. C. & Nebreda, A. R. A link between MAP kinase and p34(cdc2)/cyclin B during oocyte maturation: p90(rsk) phosphorylates and inactivates the p34(cdc2) inhibitory kinase Myt1. *EMBO J.* **17**, 5037–5047 (1998).
- Bhatt, R. R. & Ferrell, J. E. J. The protein kinase p90 rsk as an essential mediator of cytoskeletal factor activity. *Science* **286**, 1362–1365 (1999).
- Gross, S. D., Schwab, M. S., Lewellyn, A. L. & Maller, J. L. Induction of metaphase arrest in cleaving *Xenopus* embryos by the protein kinase p90Rsk. *Science* **286**, 1365–1367 (1999).
- Treinin, L., Paterson, H. F., Hooper, S., Wilson, R. & Marshall, C. J. Activated MEK stimulates expression of AP-1 components independently of phosphatidylinositol 3-kinase (PI3-kinase) but requires a PI3-kinase signal to stimulate DNA synthesis. *Mol. Cell Biol.* **19**, 321–329 (1999).
- Minden, A. *et al.* Differential activation of ERK and JNK mitogen-activated protein kinases by Raf-1 and MEKK. *Science* **266**, 1719–1723 (1994).
- Sabapathy, K. *et al.* JNK2 is required for efficient T-cell activation and apoptosis but not for normal lymphocyte development. *Curr. Biol.* **9**, 116–125 (1999).
- Xia, Z., Dickens, M., Raingeaud, J., Davis, R. J. & Greenberg, M. E. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* **270**, 1326–1331 (1995).
- Le-Niculescu, H. *et al.* Withdrawal of survival factors results in activation of the JNK pathway in neuronal cells leading to Fas ligand induction and cell death. *Mol. Cell Biol.* **19**, 751–763 (1999).
- Yang, D. *et al.* Absence of excitotoxicity-induced apoptosis in the hippocampus of mice lacking the Jnk3 gene. *Nature* **389**, 865–870 (1997).
- Behrens, A., Sibilia, M. & Wagner, E. F. Amino-terminal phosphorylation of c-Jun regulates stress-induced apoptosis and cellular proliferation. *Nature Genet.* **21**, 326–329 (1999).
- Xu, X., Raber, J., Yang, D., Su, B. & Mucke, L. Dynamic regulation of c-Jun N-terminal kinase activity in mouse brain by environmental stimuli. *Proc. Natl Acad. Sci. USA* **94**, 12655–12660 (1997).
- Tournier, C. *et al.* Requirement of JNK for stress-induced activation of the cytochrome c-mediated death pathway. *Science* **288**, 870–874 (2000).
- Shaulian, E. *et al.* The mammalian UV response: c-Jun induction is required for exit from the p53-imposed checkpoint. *Cell* **103**, 897–907 (2000).
- Kauffmann-Zeh, A. *et al.* Suppression of c-Myc-induced apoptosis by Ras signalling through PI (3)K and PKB. *Nature* **385**, 544–548 (1997).
- Bonni, A. *et al.* Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* **286**, 1358–1362 (1999).
- Pages, G. *et al.* Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* **286**, 1374–1377 (1999).
- Alberola-Illa, J., Forbush, K. A., Seger, R., Krebs, E. G. & Perlmutter, R. M. Selective requirement for MAP kinase activation in thymocyte differentiation. *Nature* **373**, 620–623 (1995).
- Giroux, S. *et al.* Embryonic death of Mek1-deficient mice reveals a role for this kinase in angiogenesis in the labyrinthine region of the placenta. *Curr. Biol.* **9**, 369–372 (1999).
- Kuan, C. Y. *et al.* The Jnk1 and Jnk2 protein kinases are required for regional specific apoptosis during early brain development. *Neuron* **22**, 667–676 (1999).
- Sabapathy, K. *et al.* Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. *Mech. Dev.* **89**, 115–124 (1999).
- Ganiatsas, S. *et al.* SEK1 deficiency reveals mitogen-activated protein kinase cascade crossregulation and leads to abnormal hepatogenesis. *Proc. Natl Acad. Sci. USA* **95**, 6881–6886 (1998).
- Hilberg, F., Aguzzi, A., Howells, N. & Wagner, E. F. c-Jun is essential for normal mouse development and hepatogenesis. *Nature* **365**, 179–181 (1993).
- Su, B. *et al.* JNK is involved in signal integration during costimulation of T lymphocytes. *Cell* **77**, 727–736 (1994).
- Li, W., Whaley, C. D., Mondino, A. & Mueller, D. L. Blocked signal transduction to the ERK and JNK protein kinases in anergic CD4+ T cells. *Science* **271**, 1272–1276 (1996).
- Sabapathy, K. *et al.* JNK1 and JNK2 have similar and stage-dependent roles in regulating T cell apoptosis and proliferation. *J. Exp. Med.* **193**, 317–318 (2001).
- Nishina, H. *et al.* Impaired CD28-mediated interleukin 2 production and proliferation in stress kinase SAPK/ERK1 kinase (SEK1)/mitogen-activated protein kinase kinase 4 (MKK4)-deficient T lymphocytes. *J. Exp. Med.* **186**, 941–953 (1997).
- Dong, C. *et al.* Defective T cell differentiation in the absence of Jnk1. *Science* **282**, 2092–2095 (1998).
- Yang, D. D. *et al.* Differentiation of CD4(+) T cells to Th1 cells requires MAP kinase JNK2. *Immunity* **9**, 575–585 (1998).
- Dong, C. *et al.* JNK is required for effector T-cell function but not for T-cell activation. *Nature* **405**, 91–94 (2000).
- Swat, W. *et al.* SEK1/MKK4 is required for maintenance of a normal peripheral lymphoid compartment but not for lymphocyte development. *Immunity* **8**, 625–634 (1998).
- Tamura, K. *et al.* Requirement for p38 α in erythropoietin expression: A role for stress kinases in erythropoiesis. *Cell* **102**, 221–231 (2000).
- Adams, R. H. *et al.* Essential role of p38alpha MAP kinase in placental but not embryonic cardiovascular development. *Mol. Cell* **6**, 109–116 (2000).
- Lu, H. T. *et al.* Defective IL-12 production in mitogen-activated protein (MAP) kinase kinase 3 (*Mkk3*)-deficient mice. *EMBO J.* **18**, 1845–1857 (1999).

Acknowledgements

We thank M. Cobb and E. Wagner for suggestions and criticism. M.K. is an American Cancer Society Research Professor; work in his lab is supported by the NIH and State of California Cancer Research Program. L.C. is supported by a NIH postdoctoral fellowship. Our apologies to those whose work was not cited due to space limitations.

Correspondence should be addressed to M. K.
(e-mail: Karinoffice@ucsd.edu).

Synaptotagmin I functions as a calcium regulator of release probability

Rafael Fernández-Chacón^{*†‡}, Andreas Königstorfer^{‡§}, Stefan H. Gerber^{*}, Jesús García^{||}, Maria F. Matos^{*}, Charles F. Stevens[‡], Nils Brose[§], Josep Rizo^{||}, Christian Rosenmund[†] & Thomas C. Südhof^{*}

^{*} Center for Basic Neuroscience, Department of Molecular Genetics, and Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, USA

[§] Max-Planck-Institut für experimentelle Medizin, and [†] Max-Planck-Institut für biophysikalische Chemie, 37070 Göttingen, Germany

^{||} Departments of Biochemistry and Pharmacology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

[‡] The Salk Institute, and Howard Hughes Medical Institute, La Jolla, California 92037, USA

[‡] These authors contributed equally to this work

In all synapses, Ca^{2+} triggers neurotransmitter release to initiate signal transmission. Ca^{2+} presumably acts by activating synaptic Ca^{2+} sensors, but the nature of these sensors—which are the gatekeepers to neurotransmission—remains unclear. One of the candidate Ca^{2+} sensors in release is the synaptic Ca^{2+} -binding protein synaptotagmin I. Here we have studied a point mutation in synaptotagmin I that causes a twofold decrease in overall Ca^{2+} affinity without inducing structural or conformational changes. When introduced by homologous recombination into the endogenous *synaptotagmin I* gene in mice, this point mutation decreases the Ca^{2+} sensitivity of neurotransmitter release twofold, but does not alter spontaneous release or the size of the readily releasable pool of neurotransmitters. Therefore, Ca^{2+} binding to synaptotagmin I participates in triggering neurotransmitter release at the synapse.

Neurotransmitter release is triggered rapidly by Ca^{2+} (in less than 1 ms)^{1–3}, but it is unknown how Ca^{2+} acts, and the identity of the Ca^{2+} sensor(s) involved is controversial^{4–7}. One of the candidates for a Ca^{2+} sensor in release is synaptotagmin I, a synaptic vesicle protein that is selectively required for fast Ca^{2+} -dependent release^{8,9}. Synaptotagmin I contains two C_2 domains (the C_2A and C_2B domains), which are composed of eight-stranded β -sandwiches, with three flexible loops emerging on top and four at the bottom^{10–13}. Both C_2 domains bind several Ca^{2+} ions exclusively through their top loops, but only Ca^{2+} binding to the C_2A domain has been studied in detail^{12,13}. The C_2A domain ligates three Ca^{2+} ions in incomplete coordination spheres formed primarily by aspartate residues in the top loops (Fig. 1).

The intrinsic Ca^{2+} affinity of the C_2A domain is low, probably because of the incomplete coordination spheres. Phospholipid membranes increase the overall Ca^{2+} affinity of the C_2A domain up to 5,000-fold (dissociation constant, $K_d \approx 10 \mu\text{M}$)^{14,15}. Native full-length synaptotagmin I also exhibits a low intrinsic Ca^{2+} -binding affinity that is increased markedly by phospholipid membranes¹⁶. Phospholipid membranes probably enhance Ca^{2+} binding to synaptotagmin I by providing additional ligands for Ca^{2+} in the incomplete coordination spheres of the Ca^{2+} -binding sites¹⁵, as was also observed for the C_2 domain of protein kinase C (PKC)^{17,18}. Thus, in the presence of Ca^{2+} and phospholipids, the C_2A domain forms a ternary complex in which Ca^{2+} is simultaneously ligated by the top loops of the C_2A domain and by phospholipid headgroups, and the phospholipids are bound by the Ca^{2+} ions and by positively charged residues that surround the Ca^{2+} -binding sites in the C_2A domain (Fig. 1).

In addition to phospholipid membranes, synaptotagmin I binds to syntaxin as a function of Ca^{2+} , although higher Ca^{2+} concentrations are required^{19–21}. At least at some synapses, release is activated at $\sim 10 \mu\text{M}$ free Ca^{2+} (refs 22–24), indicating that the exocytotic Ca^{2+} sensor has a substantially higher affinity than previously thought. The Ca^{2+} affinity corresponds well to the Ca^{2+} affinity of synaptotagmin I in the presence of phospholipids, suggesting that Ca^{2+} binding to synaptotagmin I may regulate exocytosis *in vivo*. However, confirming this function has proved technically daunting. Simply mutating Ca^{2+} -coordinating residues in a single C_2 domain

of synaptotagmin I is not effective because other Ca^{2+} -binding sites remain^{12,15}, and because interactions between the two C_2 domains²⁵ might provide additional coordination sites for Ca^{2+} . Furthermore, extensive mutations to completely abolish Ca^{2+} binding may cause unintended effects, such as mistargeting, destabilization or misfolding.

To circumvent these problems, here we have pursued an alternative approach, and studied a point mutation in synaptotagmin I that changes the Ca^{2+} -binding affinity of synaptotagmin I without affecting its three-dimensional structure. We have correlated the change in Ca^{2+} binding with the physiological consequences of the same mutation after it was introduced into the endogenous

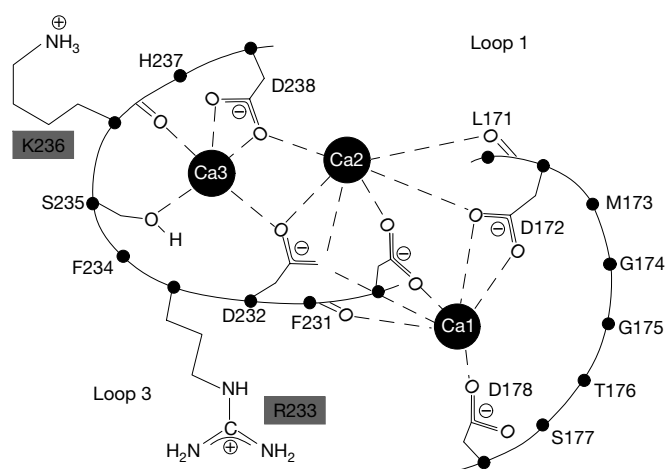


Figure 1 Architecture of the Ca^{2+} -binding sites at the top of the C_2A domain of synaptotagmin I. The three Ca^{2+} -binding sites are formed by two loops 1 and 3 at the top of the C_2A domain. Five aspartate residues, one serine residue, and two backbone carbonyl groups coordinate the three bound Ca^{2+} ions, which are labelled Ca1, Ca2 and Ca3. Arginine 233 (R233) and lysine 236 (K236), positively charged residues that surround the Ca^{2+} -binding sites and that were mutated in this study, are highlighted by a grey box. Residues are shown in single-letter amino acid code, and are identified by number (adapted from ref. 12).

synaptotagmin I gene in mice. Our results show that a mutation in synaptotagmin I that selectively decreases its apparent Ca^{2+} affinity by a factor of two causes a corresponding decrease in the Ca^{2+} sensitivity of neurotransmitter release, suggesting that Ca^{2+} binding to synaptotagmin I directly determines Ca^{2+} triggering of exocytosis.

A mutant synaptotagmin I with reduced Ca^{2+} affinity

Previous results suggested that mutations in the positively charged residues surrounding the Ca^{2+} -binding sites of the C_2A domain of synaptotagmin I influence Ca^{2+} binding^{15,26}. We therefore characterized mutations in two of these positively charged residues, R233Q and K236Q (Fig. 1), by NMR spectroscopy of the purified C_2A domains with and without Ca^{2+} . Comparison of their Ca^{2+} -free ^1H - ^{15}N heteronuclear single-quantum correlation (HSQC) spectra (Fig. 2a–c, red contours) showed that the mutations caused little perturbation in the spectra, and thus had no significant effect on the three-dimensional structure of the domain. This is not unexpected as the R233 and K236 side chains are freely exposed on the surface of the C_2A domain¹¹, and each mutation only neutralizes one of four basic residues surrounding the Ca^{2+} -binding sites (Fig. 1).

We next examined the intrinsic Ca^{2+} -binding properties of the mutants using Ca^{2+} titrations monitored with ^1H - ^{15}N HSQC spectra (Fig. 2a–c, black contours). We used Ca^{2+} -induced chemical-shift changes of cross-peaks from residues close to the Ca^{2+} -binding sites as indicators of Ca^{2+} binding. The sequential coordination of the three Ca^{2+} ions is manifested by three components in the Ca^{2+} -dependent chemical-shift changes (for example, the cross-peak of S235 HN; Fig. 2d–f, blue contours)¹². The affinities of the different Ca^{2+} -binding sites were determined from plots of the Ca^{2+} -dependent changes in chemical shifts that are strongly affected by Ca^{2+} binding to a particular Ca^{2+} -binding site. Thus, the ^{15}N chemical shift of the K200 HN cross-peak is primarily affected by Ca^{2+} binding to site Ca1 (Fig. 2a–c, arrow), and the ^1H chemical shift of the S235 HN cross-peak by Ca^{2+} binding to site Ca2 (Fig. 2d–f). Although the S235 side chain coordinates the Ca^{2+} ion in site Ca3, the S235 HN cross-peak is strongly influenced by Ca^{2+} binding to Ca2 because the peptide backbone of S235 is very close to this site¹².

In the wild-type C_2A domain, our Ca^{2+} titrations confirmed previous estimates^{12,13} of the intrinsic Ca^{2+} affinities for Ca1 ($K_d = 54 \mu\text{M}$ Ca^{2+}) and Ca2 ($K_d = 530 \mu\text{M}$ Ca^{2+}). We could not saturate binding to Ca3 even at 40 mM Ca^{2+} , indicating a K_d of more than 20 mM. The R233Q and K236Q mutants retained Ca^{2+} binding to all three sites, as expected from the lack of a structural change; however, the R233Q mutant exhibited a selectively higher Ca^{2+} affinity at site Ca1 with a K_d of 27 μM (Fig. 2), whereas sites Ca2 ($K_d = 420 \mu\text{M}$) and Ca3 ($K_d > 20 \text{ mM}$) behaved similarly to those of the wild type (Table 1). The selectively lower K_d for Ca1 in the R233Q mutant was confirmed in independent titrations with a lower protein concentration (data not shown).

The K236Q mutant of the C_2A domain exhibited Ca^{2+} affinities similar to the wild-type domain ($K_d \approx 51 \mu\text{M}$ and $K_d \approx 530 \mu\text{M}$ Ca^{2+} for Ca1 and Ca2, respectively). Ca3 may have a slightly higher affinity in the K236Q mutant, but Ca^{2+} binding to this site could not be saturated, preventing a quantitative comparison. Thus, two substitutions in positively charged residues close to the Ca^{2+} -binding sites, R233Q and K236Q, had no effect on the three-dimensional structure of the C_2A domain, but the R233Q substitution caused a selective increase in the intrinsic Ca^{2+} affinity of site Ca1, whereas the K236Q mutant did not.

R233Q decreases the overall Ca^{2+} affinity of the C_2A domain

Phospholipid membranes are thought to increase the Ca^{2+} affinity of synaptotagmin I by providing additional coordination sites for Ca^{2+} in the C_2A domain, and possibly also in the C_2B domain^{10–16}. As a result, the overall Ca^{2+} affinity of synaptotagmin I, defined here as the Ca^{2+} affinity in the presence of phospholipid membranes, is determined by formation of the entire C_2 domain/ Ca^{2+} /phospholipid complex. Thus, changes in the intrinsic Ca^{2+} affinity of a C_2 domain, such as the increased affinity for Ca1 in the R233Q mutant, may not necessarily translate into similar changes in overall Ca^{2+} affinity. We therefore measured the overall Ca^{2+} affinity of R233Q and K236Q mutant and wild-type C_2A domains using Ca^{2+} -dependent phospholipid binding as an assay with liposomes comprising 30% phosphatidylserine and 70% phosphatidylcholine (Fig. 3). As judged by this assay, the R233Q mutation caused a

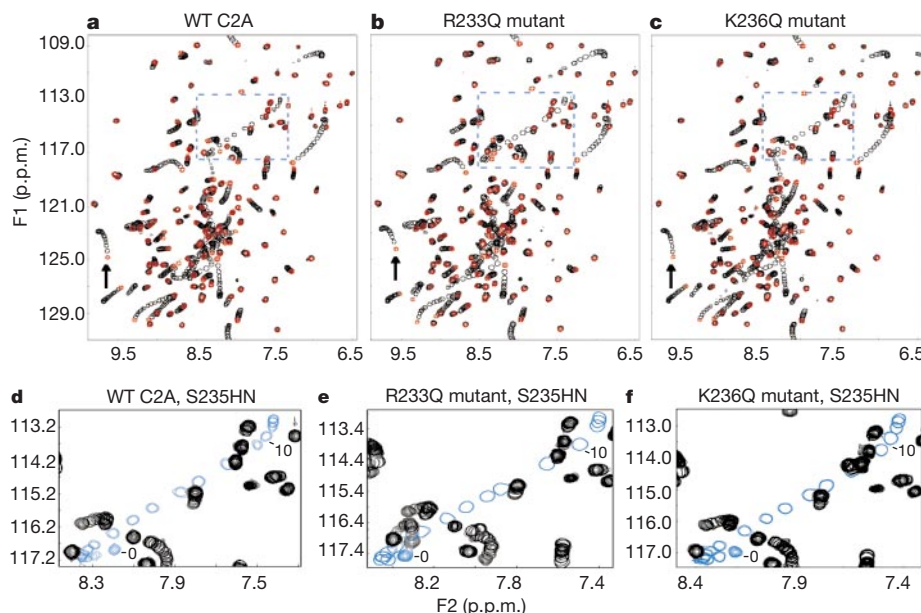


Figure 2 Ca^{2+} titrations of wild-type and mutant C_2A domains monitored by ^1H - ^{15}N HSQC spectra. **a–c**, Superposition of ^1H - ^{15}N HSQC spectra acquired at 0–40 mM Ca^{2+} concentrations. Spectra recorded in the absence of Ca^{2+} are shown in red, and spectra obtained at the various Ca^{2+} concentrations in black. Dashed boxes highlight the locations of the expansions shown in **d–f**; blue arrows identify the positions of the K200 HN cross-

peak. **d–f**, Expansions of the ^1H - ^{15}N HSQC spectra shown in **a–c** to illustrate the changes in the S235 HN cross-peak (blue contours) induced by Ca^{2+} . All other cross-peaks are shown in black. Note that titration of the S235 backbone cross-peaks (S235 HN) exhibits the largest changes on Ca^{2+} binding to site Ca2 because of the proximity of the S235 HN group to this site.

twofold decrease in overall Ca^{2+} affinity, in spite of an increase in intrinsic Ca^{2+} affinity. By contrast, the K236Q mutation had no effect on overall Ca^{2+} affinity similar to its lack of an effect on intrinsic Ca^{2+} affinity.

Formation of the C_2A domain/ Ca^{2+} /phospholipid complex should depend on the phospholipid composition, and thus the overall Ca^{2+} affinity should also be sensitive to phospholipid composition. To test this, we measured the Ca^{2+} affinity of wild-type and R233Q mutant C_2A domains with membranes containing 22.5%, 30% or 45% phosphatidylserine. As predicted, the overall Ca^{2+} affinity of the wild-type C_2A domain strongly depended on the density of negative charges on the liposome surface (Table 1). The Ca^{2+} affinity of the R233Q mutant was roughly twofold lower than that of the wild-type C_2A domain with 22.5 and 30% phosphatidylserine, plausible concentrations of negatively charged phospholipids as compared with physiological membranes. An even larger decrease in Ca^{2+} affinity was observed with 45% phosphatidylserine (Table 1).

The Ca^{2+} affinity of the C_2A domain was independent of the density of recombinant protein on the glutathione beads used in the assay because a threefold variation of this density had no effect on apparent Ca^{2+} affinity (Table 1). These results suggest that although the intrinsic Ca^{2+} affinity is enhanced in the R233Q mutant (Fig. 2), its overall Ca^{2+} affinity is depressed because phospholipids are less able to bind to the Ca^{2+} / C_2A domain complex. Surprisingly, the K236Q mutation does not have the same effect on Ca^{2+} binding as the R233Q mutation, although it is located next to R233 close to the Ca^{2+} -binding sites. This suggests that it is not the proximity of a positively charged residue to the Ca^{2+} -binding sites, but the precise location of this charge, that is critical in determining the Ca^{2+} affinity.

Generation of mice with point mutations in synaptotagmin I

We isolated a mouse genomic clone that includes a single exon encoding residues 214 to 269 of synaptotagmin I, and used this to construct three targeting vectors with the same overall design (Fig. 4)²⁷. A *neomycin* resistance gene was placed into the intron next to the exon for positive selection, and was flanked by a short arm with the exon on one side, and a long arm with intron

sequences on the other side. In addition, two copies of the thymidine kinase gene were situated at the end of the short arm for negative selection (Fig. 4). The three targeting vectors differed only in the sequence of the exon which included either the R233Q substitution, the K236Q substitution, or no substitution (to generate wild-type control mice with the same *neomycin* resistance gene in the intron). The three vectors were used for homologous recombination in embryonic stem cells, yielding mutant cell clones that were injected into blastocysts to obtain 'knockin' mice.

We produced three knockin mouse lines. All lines contain a *neomycin* resistance gene in the intron, but differ in the sequences of the adjacent exon which is either wild type or includes the R233Q or K236Q substitution. By crossing R233Q knockin and K236Q knockin mice with wild-type knockin mice, we obtained double heterozygous mice with precisely controlled genetic backgrounds. The double heterozygotes were bred to each other, resulting in homozygous wild-type or mutant knockin littermates with identical intron sequences. Both types of homozygous mutant knockin mice were viable and fertile, and exhibited no obvious morbidity. Synaptotagmin I was fully expressed in all three knockin mouse lines, indicating that the inserted *neomycin* resistance gene in the intron did not alter transcription of synaptotagmin I, and that the mutations introduced in the C_2A domain did not produce unstable or mistargeted synaptotagmin I protein (Fig. 4).

Ca^{2+} affinity of native wild-type and mutant synaptotagmin I

To confirm that the R233Q substitution affects Ca^{2+} binding to native brain synaptotagmin I and not only to the isolated C_2A domain, we examined the Ca^{2+} -dependent phospholipid- and syntaxin-binding properties of native wild-type and R233Q mutant synaptotagmin I. Full-length synaptotagmin I is prone to nonspecific interactions, presumably because its amino-terminal transmembrane region is highly palmitoylated and indiscriminately binds to hydrophobic surfaces. Thus, to avoid nonspecific interactions, we removed the hydrophobic N-terminal region of synaptotagmin I by cleaving synaptotagmin I at the hypersensitive

Table 1 Ca^{2+} affinities of wild-type and mutant synaptotagmins

Condition	Protein		
	Wild type	R233Q	K236Q
Intrinsic Ca^{2+} affinity in the absence of phospholipids*			
Ca^{2+} -binding site Ca1	54 μM	27 μM	51 μM
Ca^{2+} -binding site Ca2	530 μM	420 μM	530 μM
Ca^{2+} -binding site Ca3	>10 mM	>10 mM	>10 mM
Overall Ca^{2+} affinity in the presence of phospholipid membranes†			
22.5% PS/77.5% PC	40.8 $\pm$ 1.9 μM	85.0 $\pm$ 7.1 μM	n.d.
30.0% PS/70.0% PC (high density)	11.1 $\pm$ 3.0 μM	23.9 $\pm$ 4.6 μM	10.6 $\pm$ 1.7 μM
30.0% PS/70.0% PC (low density)	12.7 $\pm$ 1.3 μM	29.7 $\pm$ 4.5 μM	n.d.
45.0% PS/55.0% PC	4.1 $\pm$ 0.2 μM	24.5 $\pm$ 6.0 μM	n.d.
Overall Ca^{2+} affinity of native synaptotagmin C_2 domains‡			
30.0% PS/70.0% PC	~3–4 μM	~6–8 μM	n.d.

* Dissociation constants were calculated from Ca^{2+} titrations monitored by ^1H - ^{15}N HSQC NMR spectra acquired with 150–160 μM recombinant C_2A domains. The Ca^{2+} dependence of the ^{15}N chemical shift of K200 NH resonances from 0.0–0.6 mM total Ca^{2+} , and of the ^1H chemical shift of the S235 NH resonance from 0.5–2.0 mM total Ca^{2+} were used for Ca^{2+} -binding sites Ca1 and Ca2, respectively. Dissociation constants for Ca^{2+} -binding site Ca3 were estimated as it could not be reliably calculated because the binding site was never fully saturated.

† Binding constants represent the EC_{50} obtained from Ca^{2+} titrations of the binding of radiolabelled liposomes to GST- C_2A domain fusion proteins (Fig. 3). Data are means $\pm$ s.e.m. from 3–9 independent experiments performed in triplicate with liposomes comprising the indicated phosphatidylserine (PS)/phosphatidylcholine (PC) mixes (n.d. = not determined). For 30% PS/70% PC liposomes, beads were either saturated with the fusion protein (high density), or the fusion protein was diluted in a 1:2 ratio with GST before binding to the beads (low density) to test the effect of protein density on the beads on the apparent Ca^{2+} affinity.

‡ Determined by binding of synaptotagmin cytoplasmic fragments to liposomes using a centrifugation assay (Fig. 5).

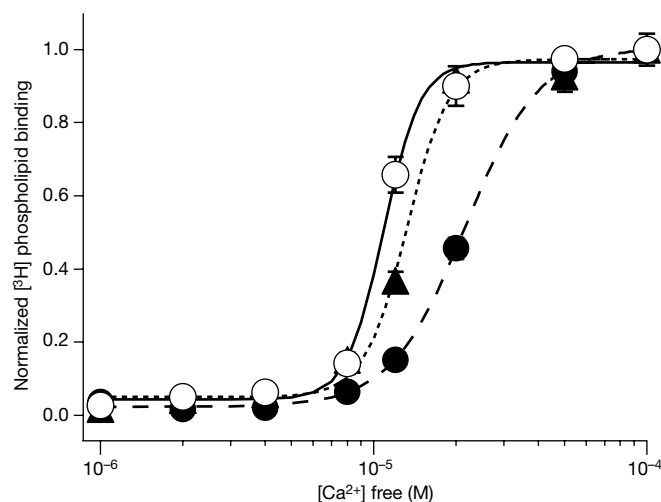


Figure 3 Ca^{2+} dependence of phospholipid binding to wild-type and mutant C_2A domains. Liposomes composed of 30% phosphatidylserine/70% phosphatidylcholine were bound to immobilized GST fusion proteins of wild-type (open circles), R233Q mutant (filled circles) and K236Q mutant (triangles) C_2A domains from synaptotagmin I at the indicated concentrations of free Ca^{2+} clamped with Ca^{2+} /EGTA buffers. Graphs show the means $\pm$ s.e.m. of a representative experiment performed in triplicate, and normalized to the binding at saturating Ca^{2+} concentrations; most error bars are not visible because of their small size. EC_{50} values for Ca^{2+} : wild-type, 11 μM ; K236Q, 13 μM ; R233Q, 21 μM . Results from several independent experiments with different lipid compositions are summarized in Table 1.

proteolytic site between the transmembrane region and the C₂A domain^{28,29}.

We briefly treated brain homogenates from wild-type and R233Q mutant brains with trypsin, and isolated the cleaved C₂A/C₂B domain fragment in the supernatant after pelleting the membranes. The native C₂A/C₂B domain fragment was used in Ca²⁺-dependent phospholipid-binding assays by incubating it with liposomes in the presence of defined concentrations of free Ca²⁺. Liposomes were pelleted, and the amount of C₂A/C₂B domain fragment bound was determined by immunoblotting (Fig. 5a). In parallel, we incubated the C₂A/C₂B domain fragment with glutathione beads containing either glutathione *S*-transferase (GST)–syntaxin or GST alone at different Ca²⁺ concentrations, and measured the amount of bound C₂A/C₂B domain fragment by immunoblotting of the pelleted beads (Fig. 5b).

The data show that the R233Q mutant C₂A/C₂B domain fragment bound phospholipids and syntaxin to a similar extent as did the wild-type C₂A/C₂B domain fragment. However, the R233Q mutant C₂A/C₂B domain fragment exhibited a roughly twofold lower overall Ca²⁺ affinity than that of the wild-type fragment (Fig. 5a), similar to the results obtained with recombinant C₂A domains (Table 1). We observed no consistent difference in syntaxin binding between wild-type and mutant C₂A/C₂B domains, possibly because the more labile, syntaxin-binding assay makes it difficult to detect a twofold shift in Ca²⁺ affinity (Fig. 5b). Notably, the native C₂A/C₂B domain fragment bound to syntaxin at much lower Ca²⁺ concentrations (< 20 μM) than previously described for recombinant fragments (> 200 μM Ca²⁺)^{19,21}. This result suggests that in native synaptotagmin I, the two C₂ domains cooperate with each other in

Ca²⁺ binding and/or are post-translationally modified. Such cooperation and/or modifications could also explain why our previous results suggested that the R233Q mutant C₂A domain is impaired in syntaxin binding²⁶, whereas the native double C₂-domain fragment is not.

R233Q selectively impairs neurotransmitter release

To test the effects of the R233Q and K236Q mutations on synaptotagmin I function, we cultured hippocampal neurons from newborn knockin mice, and performed electrophysiological recordings in neurons that had formed autapses^{9,30–32}. In all experiments, neurons from mutant and wild-type littermates were cultured and analysed simultaneously to control for the variability in synaptic responses between cultures. Neurons from R233Q, K236Q and wild-type knockin mice exhibited robust excitatory postsynaptic currents (EPSCs) in response to action potentials induced by brief depolarizations. However, we observed marked differences between the mutants in the relative EPSC amplitudes during repetitive stimulation (Fig. 6). The EPSC amplitudes of R233Q mutant neurons were smaller than those of wild-type neurons (see below), but displayed large facilitation upon stimulation at 10 Hz, 20 Hz or 50 Hz. Under the same conditions, both wild-type knockin neurons and K236Q knockin neurons exhibited mild synaptic depression (Fig. 6). As a result, the relative synaptic response at the end of a stimulus train was several-fold higher in R233Q mutant neurons than in wild-type neurons and in K236Q mutant neurons,

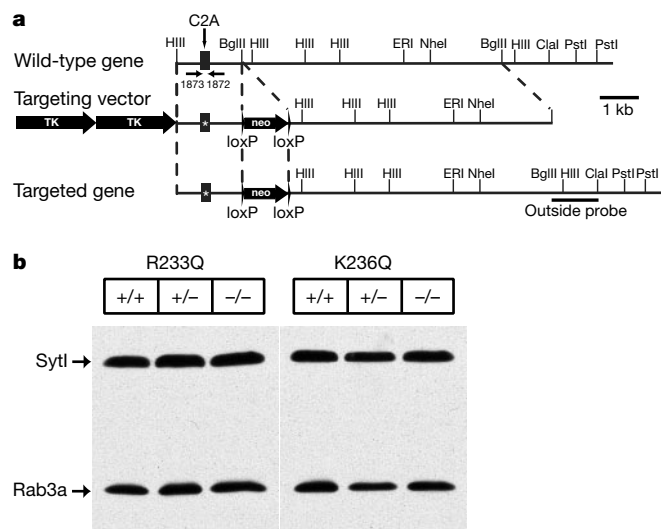


Figure 4 Generation of mice with mutant C₂A domains. **a**, Strategy for creating knockin mice by homologous recombination. A genomic clone containing a single exon from the murine synaptotagmin I gene (top) was used to construct three targeting vectors of the same overall design (middle), with two copies of the thymidine kinase (TK) gene for negative selection and a *neomycin* resistance cassette (*neo*) for positive selection. In the three vectors, the exon contains the R233Q or the K236Q substitution (asterisk) or is wild type (to generate a precisely matched wild-type control). Homologous recombination replaces the endogenous synaptotagmin I exon with the exon from the targeting vector, and inserts the *neomycin* resistance gene cassette into the intron (bottom). Numbered arrows identify the oligonucleotides (1,872 and 1,873) used for genotyping. The position of the outside probe for detection of homologous recombination by Southern blotting is indicated on the right. Locations of selected restriction sites are shown (HIII = *HindIII*; ERI = *EcoRI*), and the scale is given on the right. **b**, Immunoblots of total brain homogenates (30 μg protein) from wild-type (+/+), heterozygous (+/-) and homozygous (-/-) littermate mice stained with monoclonal antibodies specific for synaptotagmin I and rab3a.

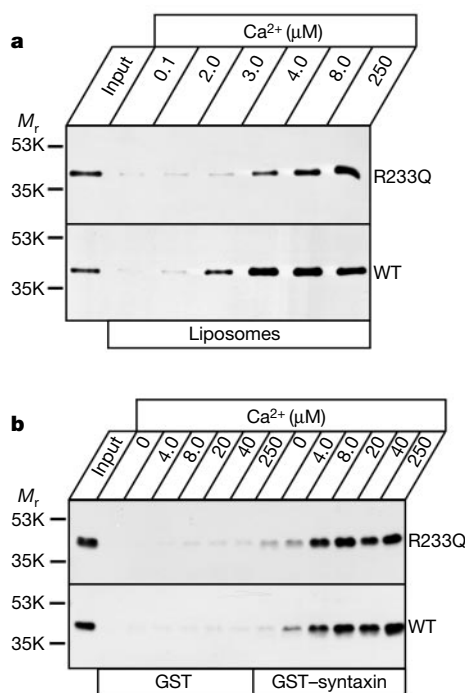


Figure 5 Effect of the R233Q mutation on Ca²⁺-dependent phospholipid and syntaxin binding by native synaptotagmin. **a**, The soluble C₂A/C₂B-domain fragment of synaptotagmin I was obtained from R233Q mutant and wild-type (WT) mice by partial trypsin digestion of synaptotagmin I in brain homogenates. The soluble C₂A/C₂B-domain fragment isolated in the supernatant of centrifuged trypsinized brain homogenates was used for binding experiments with liposomes at the indicated concentrations of free Ca²⁺. The input and the proteins bound to the liposomes at the various Ca²⁺ concentrations were analysed by immunoblotting using a monoclonal synaptotagmin I antibody. Numbers on the left indicate positions of size markers. **b**, Pull-down experiments of the soluble C₂A/C₂B-domain fragment of synaptotagmin I isolated as described above with GST and with GST-syntaxin (residues 180–264). Experiments were performed with the indicated concentrations of free Ca²⁺ as described in **a**, and bound proteins were visualized by immunoblotting.

suggesting a selective functional effect of the R233Q substitution.

Facilitation during repetitive stimulation is indicative of a reduced release probability^{33,34}, suggesting that the R233Q mutation changes the efficacy with which Ca^{2+} triggers release during an action potential. The synaptic release probability is the probability with which an action potential stimulates neurotransmitter release at a synapse. It can be measured as the rate of inhibition by an irreversible open-channel blocker of the NMDA (*N*-methyl-D-aspartate) receptor, MK-801 (refs 35, 36). MK-801 only blocks NMDA receptors that have been opened by glutamate, thereby 'knocking out' synapses as they become activated.

To evaluate the effect of the R233Q mutation on synaptic release probability, we measured the rate of inhibition of NMDA-receptor-dependent EPSCs by MK-801. In R233Q mutant and in wild-type neurons, MK-801 caused a multi-exponential decay in the EPSC amplitude, which is consistent with two or more populations of synapses having distinct release probabilities^{35,36} (Fig. 7). The decay rate was significantly lower in R233Q mutant neurons than in wild-type neurons, however, showing that the R233Q substitution causes a reduction in synaptic release probability. Assuming that the distribution of release probabilities among synapses is similar between wild-type and R233Q mutant neurons, the decrease in

release probability can be quantified as the degree of expansion in the *x* axis of the wild-type curve that is necessary to make it fit to the mutant curve (Fig. 7, inset). This procedure showed that a twofold increase in the scale of the wild-type *x* axis induced a perfect match to the mutant response, suggesting that the R233Q mutation decreased the synaptic release probability twofold.

Vesicular release probability in R233Q mutant mice

A reduction in synaptic release probability could be caused by a decline in the readily releasable pool of synaptic vesicles or by a decrease in the Ca^{2+} triggering of synaptic vesicle exocytosis. In the first case, fewer vesicles would be readily available for release per synapse, but would be normally Ca^{2+} responsive. In the second case, the number of readily releasable vesicles would be unchanged, but the effectiveness of Ca^{2+} influx during an action potential to trigger release would be impaired. To differentiate between these possibilities, we measured consecutively the amplitude of the Ca^{2+} -triggered synaptic response and the size of the readily releasable pool of vesicles in the same neuron (Fig. 8a).

The amplitude of Ca^{2+} -triggered synaptic responses was determined by inducing action potentials with short pulses of depolarization. Evoked responses were then recorded from a large number of mutant and control neurons to obtain representative average EPSC amplitudes. Consistent with the decrease in synaptic release probability (Fig. 7), the amplitude of evoked EPSCs was significantly reduced in R233Q mutant neurons (to ~55 % of the wild-type response) (Fig. 8b). No amplitude reduction was found in K236Q mutant neurons, similar to the lack of a phenotype in the repetitive stimulation paradigm.

To measure the size of the readily releasable pool of synaptic vesicles, we applied hypertonic sucrose (0.5 M), which stimulates exocytosis of the entire readily releasable pool of vesicles by an unknown, Ca^{2+} -independent mechanism^{37,38}. In contrast to Ca^{2+} -evoked EPSC amplitudes, no difference in the size of the sucrose

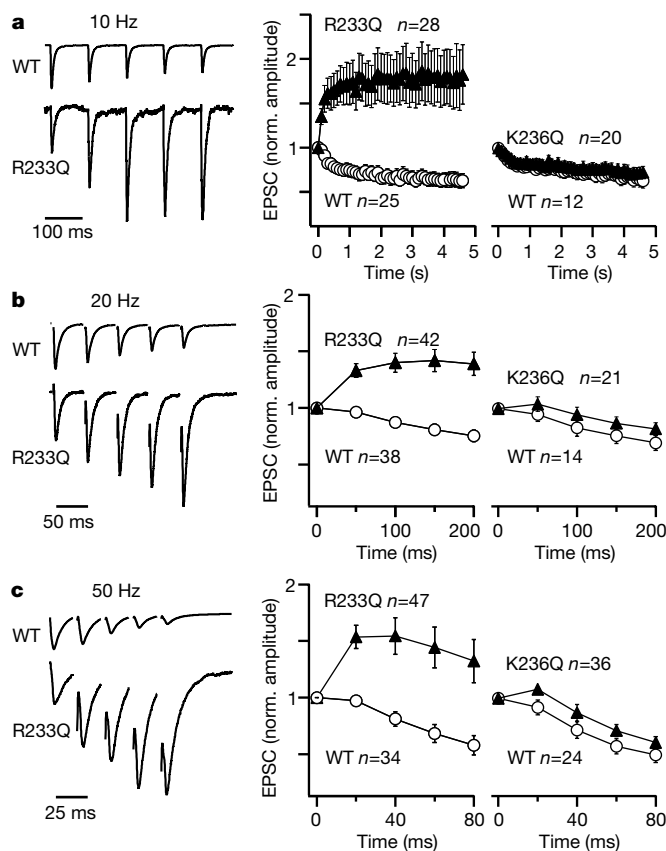


Figure 6 Effect of the R233Q and K236Q mutations on synaptic responses in cultured hippocampal neurons. Evoked synaptic responses were recorded in autapses of cultured hippocampal neurons stimulated at 10 Hz (**a**), 20 Hz (**b**) and 50 Hz (**c**). For each stimulation frequency, representative traces from wild-type (WT) and R233Q mutant neurons are shown on the left, and summary graphs for the respective knockin mutants are shown middle and right for R233Q and K236Q, respectively. Summary graphs compare individual signals from multiple experiments in which mutant and wild-type knockin neurons from littermates were analysed simultaneously to control for culture conditions (means $\pm$ s.e.m.). Recordings were made in standard 4 mM Ca^{2+} , 4 mM Mg^{2+} solutions. Vertical calibration bars have been omitted since the currents are normalized to the first response because EPSC amplitudes are lower in the R233Q knockin neurons than in wild-type knockin or K236Q knockin neurons (see Fig. 8).

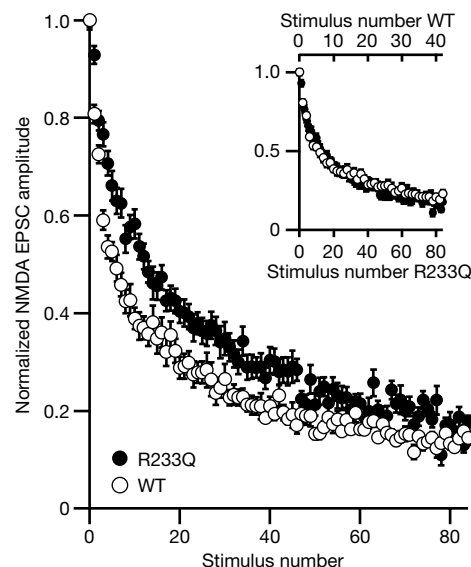


Figure 7 Synaptic release probability in hippocampal neurons from R233Q knockin and wild-type knockin mice. The rate of blockade of excitatory synaptic transmission was recorded as a function of stimulus number in the presence of the irreversible NMDA-receptor inhibitor MK-801 (5 μM). The responses of mutant and wild-type neurons for the same stimulus numbers are compared; the inset shows a plot of the wild-type and mutant signals in which the *x* axis was scaled differentially by a factor of two to match the curves. Responses were recorded in 2.7 mM Ca^{2+} , 0 mM Mg^{2+} , and normalized to the initial synaptic amplitude in MK-801 (~70% of the control NMDA-receptor signal). Data shown are means $\pm$ s.e.m. ($n = 9$ and 16 for wild-type and R233Q mutant neurons, respectively).

response was detected between R233Q mutant neurons and K236Q mutant or wild-type neurons (Fig. 8b). Thus, the R233Q substitution in synaptotagmin I caused a selective decrease in evoked EPSCs that represent Ca^{2+} -triggered fusion, but had no effect on the sucrose response that represents the readily releasable pool. In further support of this conclusion, the frequency and amplitude of spontaneous miniature postsynaptic currents (mEPSCs) were similar in R233Q mutant and wild-type neurons (wild-type, 1.9 ± 0.8 events per second with 13.7 ± 2.5 pA, $n = 5$; R233Q mutant, 2.5 ± 0.8 events per second with 11.5 ± 1.5 pA, $n = 6$).

The vesicular release probability is the probability with which a particular vesicle in the readily releasable pool can be stimulated for exocytosis by Ca^{2+} influx during an action potential^{32,38}. This probability can be calculated as the ratio of Ca^{2+} -evoked release to the size of the readily releasable pool. Although in cultured autaptic neurons the sizes of the EPSC amplitudes and of the readily releasable pool vary between cultures, the vesicular release probability of a single neuron is reliably determined when the synaptic responses to action potentials and hypertonic sucrose are recorded in the same neuron (Fig. 8a). For each form of release, the measured charge integral of the postsynaptic currents is the product of the charge of individual mEPSCs (corresponding to release of single vesicles) and the number of released vesicles. Thus, the ratio of the charge integrals of Ca^{2+} -evoked and sucrose-stimulated release measured in a single cell directly corresponds to its vesicular release probability.

When we performed such calculations for the data shown in Fig. 8b, we observed an almost threefold decrease in the vesicular release probability in R233Q mutant neurons compared with that in wild-type knockin neurons that were analysed simultaneously ($2.34 \pm 0.32\%$, $n = 44$, versus $6.95 \pm 0.85\%$, $n = 44$; Fig. 8c). No significant difference was detected in the vesicular release probability between K236Q mutant and littermate wild-type neurons ($4.6 \pm 0.7\%$, $n = 32$, versus $5.2 \pm 0.9\%$, $n = 23$). Although substantial, the threefold reduction in vesicular release probability in the R233Q mutants is smaller than the tenfold reduction in the release probability observed under the same conditions in synaptotagmin I knockouts ($0.6 \pm 0.2\%$, $n = 12$; data not shown). The reduction in vesicular release probability in the R233Q mutants (Fig. 8c) is slightly higher than the reduction in synaptic release probability measured by the rate of MK-801 blockade (Fig. 7), probably because of the distinct Ca^{2+} and Mg^{2+} concentrations that were required for these experiments.

Decreased exocytotic Ca^{2+} sensitivity in R233Q synapses

To obtain a quantitative estimate for the relative Ca^{2+} sensitivity of exocytosis in R233Q mutant and wild-type neurons, we monitored the size of evoked EPSCs as a function of the external Ca^{2+} concentration. Evoked EPSCs were recorded during low-frequency stimulation (0.2 Hz for 30–60 s) at variable Ca^{2+} concentrations (1–12 mM) with a constant Mg^{2+} concentration (1 mM Mg^{2+}) to maximize the release probability, and to allow examination of a large range of Ca^{2+} concentrations. Each test measurement was followed by a control measurement in standard 4 mM Ca^{2+} , 4 mM Mg^{2+} solution to correct for the rundown of synaptic responses during the experiment. We then normalized the response at each test Ca^{2+} concentration to the response at the standard $\text{Ca}^{2+}/\text{Mg}^{2+}$ concentration obtained immediately afterwards. As the absolute magnitude of the EPSC amplitude is significantly decreased in the R233Q mutant (Fig. 8b), we also normalized the mutant signal to the wild-type signal under standard conditions. In this way, we measured the Ca^{2+} dependence of release independently of the absolute EPSC amplitude.

The resulting Ca^{2+} -response curves reveal two features (Fig. 8d). First, synaptic amplitudes in the R233Q mutant neurons are consistently lower than in wild-type neurons, as expected from the relative depression of release under standard conditions in this

mutant (see Fig. 8b). Second, the Ca^{2+} -response curve in the mutant is shifted to the right compared with the wild-type Ca^{2+} -response curve. In the mutant neurons, we could not achieve saturation because increases in extracellular Ca^{2+} beyond 12 mM led to action potential failures and to saturation of Ca^{2+} influx³⁹. Nevertheless, high Ca^{2+} concentrations at least partially rescued the mutant phenotype. At the highest Ca^{2+} concentration used (12 mM), the wild-type EPSC amplitude was 1.6 times the control response in standard 4 mM Ca^{2+} , 4 mM Mg^{2+} solution, whereas in R233Q mutant neurons we observed a 2.6-fold increase over the control response (Fig. 8d).

To calculate the apparent affinity for extracellular Ca^{2+} , the dose-response curve was fitted with the Hill function ($I = I_{\text{max}} / (1 + (K_d / [\text{Ca}^{2+}])^n)$), where I is the recorded synaptic current plotted as normalized amplitude, I_{max} the maximal current, K_d the apparent dissociation constant for extracellular Ca^{2+} , $[\text{Ca}^{2+}]$ the extracellular Ca^{2+} concentration, and n the apparent cooperativity. The calculated Ca^{2+} affinity was decreased twofold in R233Q mutant neurons compared to wild-type neurons ($K_d = 3.7 \pm 0.8$ mM versus 1.9 ± 0.2 mM extracellular Ca^{2+}). By contrast, the cooperativity of release, as reflected in the Hill coefficient, was similar (1.7 for wild-type neurons versus 1.6 for R233Q mutant neurons), and the calculated I_{max} of the R233Q neurons was 74% of the wild-type value (1.27 versus 1.72). Note that the external Ca^{2+} affinity and Ca^{2+} cooperativity obviously represent an underestimate of true intracellular Ca^{2+} affinities and cooperativities because they are only indirectly related³⁹.

Intracellular Ca^{2+} responsiveness in R233Q-mutant synapses

The most plausible explanation for the impaired vesicular release probability in R233Q mutant neurons is that this mutation directly affects the function of the Ca^{2+} sensor for synaptic vesicle exocytosis; however, the impaired vesicular release probability could also be explained by a decrease in Ca^{2+} influx. As synaptotagmin I and syntaxin are known to interact with Ca^{2+} channels^{40,41}, it is conceivable that the decreased Ca^{2+} responsiveness of synaptic vesicles in the R233Q mutant neurons is due to an inhibition of Ca^{2+} influx. To examine this, we bypassed Ca^{2+} -channel activation by applying the Ca^{2+} ionophore calcimycin. Calcimycin spontaneously inserts into the plasma membrane, and allows Ca^{2+} to flow directly into the neurons. Soon after application of calcimycin, robust Ca^{2+} -induced spontaneous EPSCs were observed in wild-type neurons (Fig. 8e). Application of the same concentration of calcimycin under identical conditions to R233Q mutant neurons led to a delay in the stimulation of EPSCs compared to wild-type neurons (Fig. 8e).

The speed with which a synaptic response is triggered by a Ca^{2+} ionophore will depend on how fast the ionophore inserts into the membrane, how fast Ca^{2+} flows through the ionophore into the terminal, and how much intracellular Ca^{2+} is required in the terminal to trigger release. As the same concentrations of Ca^{2+} ionophore were applied to wild-type and R233Q mutant neurons, the rise in intracellular Ca^{2+} should be identical. Thus, the delayed response in the mutant neurons suggests that higher local Ca^{2+} concentrations in the nerve terminal are necessary to trigger neurotransmitter release from R233Q mutants, consistent with a change in the Ca^{2+} responsiveness of synaptic vesicles downstream of Ca^{2+} channels.

Conclusions

We examined two point mutations in synaptotagmin I, R233Q and K236Q, that neutralize positively charged residues close to the Ca^{2+} -binding sites but have no effect on the three-dimensional structure of synaptotagmin I. R233Q reduces the overall Ca^{2+} affinity of synaptotagmin I, and results in a selective decrease in the Ca^{2+} responsiveness of synaptic vesicle exocytosis. By contrast, K236Q had no effect *in vitro* or *in vivo*. Thus a single-point mutation in synaptotagmin I selectively decreases its overall Ca^{2+} affinity by a

factor of two, and also causes a 2–3-fold decrease in the Ca^{2+} sensitivity of release. The R233Q phenotype is not due to non-specific electrostatic changes because the neighbouring K236Q mutation does not alter the Ca^{2+} affinity of synaptotagmin I or the Ca^{2+} responsiveness of release. The most parsimonious explanation for the effect of the R233Q mutation is that Ca^{2+} binding to synaptotagmin I triggers synaptic vesicle exocytosis; however, this explanation does not imply that synaptotagmin I is the only Ca^{2+} sensor in release as several Ca^{2+} sensors may have to be activated simultaneously to trigger release (for example, several synaptotagmin isoforms).

The similarity of the overall Ca^{2+} affinity of synaptotagmin^{14–16} with that of the exocytotic Ca^{2+} sensor in central synapses^{22–24}, and the parallel shift in these parameters caused by the R233Q mutation, strongly support the role of synaptotagmin I as a Ca^{2+} sensor in release. Nevertheless, alternative explanations could also account for the observed phenotype of the R233Q mutation, although they appear to be less likely than the Ca^{2+} -sensor hypothesis. One alternative explanation is that Ca^{2+} influx is inhibited, leading to a decrease in the apparent Ca^{2+} sensitivity of release. This explanation is unlikely, however, because the Ca^{2+} ionophore experiments revealed that Ca^{2+} triggering is impaired in the R233Q mutant downstream of Ca^{2+} influx.

A second potential explanation for the decrease in apparent Ca^{2+}

responsiveness is that Ca^{2+} buffering is increased. This explanation is also improbable because the R233Q mutation causes only a small change in Ca^{2+} binding compared with the large increase in Ca^{2+} buffering required to account for the phenotype. A third potential alternative explanation for the R233Q phenotype is that R233 mediates binding of an unidentified protein to synaptotagmin I which is the 'true' Ca^{2+} sensor in release instead of synaptotagmin. To be the true Ca^{2+} sensor, however, the unidentified synaptotagmin I ligand—possibly the SNARE complex—would itself have to bind Ca^{2+} but interact equally well with Ca^{2+} -bound and Ca^{2+} -free synaptotagmin I (otherwise synaptotagmin I would, after all, still be a Ca^{2+} sensor in release). These conditions make this explanation also rather unlikely, especially in view of the role of R233 in Ca^{2+} binding to synaptotagmin I.

What is the mechanism of action of synaptotagmin I as a Ca^{2+} sensor/transducer in release? In proteins such as perforin, PKC and phospholipase C, phospholipid binding to C₂ domains mediates the Ca^{2+} -dependent regulation^{42–45}. By analogy, it seems plausible that binding of the C₂ domains of synaptotagmin I to phospholipid membranes is integral to its function. Synaptotagmin I is present on docked vesicles at the active zone close to the plasma membrane, and thus has ready access to phospholipid membranes. In addition, the interaction of synaptotagmin I with syntaxin provides an attractive mechanism of action because of the fundamental role of

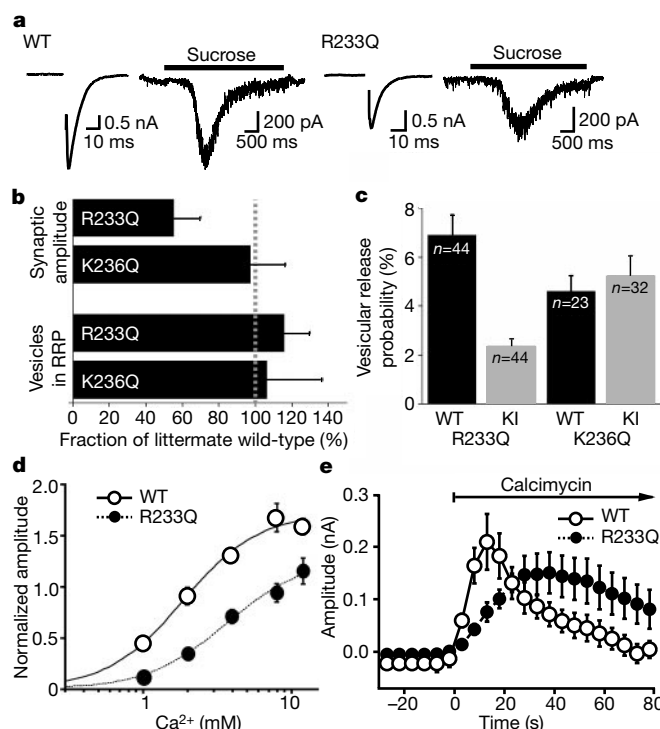


Figure 8 Pool sizes and Ca^{2+} responsiveness of synaptic vesicles in wild-type, R233Q and K236Q mutant synapses. **a**, Typical synaptic responses evoked by action potentials or application of hypertonic sucrose in neurons from wild-type and R233Q knockin mice. Action potentials were induced by short somatic depolarization (from -75 mV to 0 mV for 2 ms), and the readily releasable pool of vesicles was stimulated from the same cell by hypertonic sucrose (0.5 M for 3 s). Capacitative and somatic currents are blanked for display purposes. **b**, Quantitative analysis of the synaptic EPSC amplitudes and the size of the readily releasable pool of R233Q and K236Q knockin neurons compared with wild type. Mean amplitudes and pool sizes recorded in mutant knockin neurons were normalized to wild-type knockin littermate controls; responses are expressed as per cent of wild-type controls (100% = dotted line; n is the same as in **c**). **c**, Mean vesicular release probability expressed as a per cent of the readily releasable pool. R233Q and K236Q knockin neurons were analysed simultaneously with wild-type knockin control neurons from littermates, and are shown in comparison with these controls to correct for culture-dependent variability. Vesicular release probability was obtained from recordings in which

evoked EPSCs and the readily releasable pool were measured in the same cells, and is calculated as the ratio of the evoked EPSC charge integral to the sucrose-stimulated charge integral. **d**, Ca^{2+} dependence of release from wild-type knockin (WT) and R233Q knockin synapses. Synaptic responses to brief depolarizations were recorded at the indicated concentrations of Ca^{2+} in the presence of 1 mM Mg^{2+} . The same neurons were examined consecutively for all Ca^{2+} concentrations, with control measurements in standard 4 mM Ca^{2+} , 4 mM Mg^{2+} solutions after each test measurement to correct for the rundown of responses during recordings; responses are additionally normalized for the absolute EPSC amplitudes. Data shown are means $\pm$ s.e.m. ($n = 6$ – 16 and 8 – 18 cells for wild-type and R233Q knockin neurons, respectively). **e**, Spontaneous synaptic EPSCs evoked by the Ca^{2+} ionophore calcimycin in wild-type (WT) and R233Q knockin neurons. Calcimycin was applied at time 0 after stable baseline recordings for 100 s. Data shown are means $\pm$ s.e.m. ($n = 9$ and 12 cells for wild-type and R233Q knockin neurons, respectively). All recordings except for those in **d** were made in standard 4 mM Ca^{2+} , 4 mM Mg^{2+} solutions.

syntaxin in synaptic vesicle fusion⁴⁶. The absolute Ca^{2+} dependence and the markedly lower Ca^{2+} requirement for syntaxin binding to native synaptotagmin (as opposed to recombinant synaptotagmin^{19,21}) fit well with the role of an exocytotic Ca^{2+} sensor^{22–24}. Although we were unable to detect a change in syntaxin binding in the R233Q mutant, it is possible that Ca^{2+} -dependent phospholipid and syntaxin binding to synaptotagmin I cooperate in Ca^{2+} triggering of exocytosis. This might occur together with additional Ca^{2+} -dependent activities of the C_2 domains, such as the Ca^{2+} -dependent multimerization of synaptotagmin I (refs 47, 48). Independent of the precise mechanism of action of synaptotagmin I, however, our data show that synaptotagmin I is a direct component of the Ca^{2+} sensor in synchronous release where alterations in its overall Ca^{2+} affinity translate into equivalent changes in the Ca^{2+} sensitivity of neurotransmitter release. □

Methods

Ca^{2+} -binding studies of C_2A domains

^1H - ^{15}N HSQC spectra were acquired at 25 °C in a Varian INOVA500 NMR spectrometer (spectral widths = 7,600 and 1,163 Hz in the ^1H and ^{15}N dimensions, respectively) using uniformly ^{15}N -labelled C_2A domains^{11–13,49}. Initial Ca^{2+} titrations were recorded at 150–160 μM protein concentrations using ^1H - ^{15}N HSQC spectra (1 h total acquisition time) with successive additions of 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.375, 0.45, 0.525, 0.6, 0.675, 0.75, 1, 1.2, 2, 5, 10, 20, 30 and 40 mM Ca^{2+} . Additional Ca^{2+} titrations of 18–20 μM wild-type and R233Q mutant C_2A domains were obtained with ^1H - ^{15}N HSQC spectra (12 h total acquisition time) recorded after successive additions of 0, 10, 20, 40, 60, 80, 100, 140, 180 and 220 μM Ca^{2+} (wild-type) and 0, 5, 10, 20, 40, 60, 80 and 100 μM Ca^{2+} (mutant). The Ca^{2+} affinities of the different sites were calculated by curve fitting the Ca^{2+} dependence of selected chemical shifts to standard protein–ligand equilibrium equations using Sigma Plot.

Phospholipid-binding studies of C_2A domains

^3H -labelled liposomes containing the indicated concentrations of PS and PC were prepared by sonication, and used in phospholipid-binding assays with immobilized GST or GST fusion proteins of the C_2A domains^{14,15}. Binding assays were carried out with identical amounts of liposomes, protein and glutathione–agarose using Ca^{2+} /EGTA buffers calculated with EqCal (Biosoft, Cambridge, UK). To test the effect on the apparent Ca^{2+} affinity of the density of the C_2A domain fusion protein on the beads, beads were loaded with a 2:1 mixture of GST to the GST– C_2A domain fusion protein, and the binding assays were performed with these diluted C_2A domains.

Phospholipid and syntaxin binding

Post-nuclear supernatants from mouse brains in 20 mM HEPES–NaOH pH 7.2, 0.1 g l^{–1} PMSF were diluted to 8 g protein per l and digested with 0.008% trypsin for 30 min at 37 °C. Afterwards, 0.1 g l^{–1} soybean trypsin inhibitor, 1% goat serum, and 1 mM PMSF were added, samples were centrifuged at 300,000g for 4 h at 4 °C, and the supernatants were re-centrifuged for an additional 30 min. Aliquots (1.1 ml) of the supernatants and of 2.4× Ca^{2+} /EGTA-buffers in 80 mM HEPES–NaOH pH 7.2, 0.2 M NaCl were mixed. One millilitre of each sample was distributed into two tubes, 0.2 ml of 0.5 g l^{–1} liposomes were added in 50 mM HEPES–NaOH pH 7.2, 0.1 M NaCl, and samples were incubated at 36 °C for 15 min under agitation. Liposomes were pelleted by centrifugation (100,000g for 30 min), washed once in the appropriate Ca^{2+} /EGTA buffer, and analysed by SDS–PAGE and immunoblotting using Cl41.1 synaptotagmin I monoclonal antibody.

For GST–syntaxin binding assays, trypsinized brain extracts were aliquoted with Ca^{2+} /EGTA buffers as described above with 1% Triton X-100 in the final solutions. Agarose beads with GST–syntaxin 180–264 or GST alone (1–2 g protein per l glutathione agarose) were preblocked in 20 mM HEPES–NaOH pH 7.2, 1% BSA for 30–90 min at 4 °C, pelleted, and 1.2 ml of trypsinized brain extract in Ca^{2+} /EGTA buffer was added. Samples were incubated overnight at 4 °C on a rotator, pelleted, washed three times with the appropriate incubation buffers, and pelleted again. Bound proteins were analysed by SDS–PAGE and immunoblotting as described above.

Generation of knockin mutant mice

We used a genomic λ -clone containing a single exon with residues 214–269 of murine synaptotagmin I to construct three similar targeting vectors with either wild-type or R233Q and K236Q mutant exon sequences (Fig. 4). Vectors were electroporated into embryonic stem cells (E14 cells, a gift from K. von Figura, Göttingen), colonies were selected with G418 and gancyclovir²⁷, and double-resistant colonies were analysed for homologous recombination by Southern blotting with an outside probe. Clones containing homologously recombined genes were expanded, confirmed by PCR, and used to generate mice by blastocyst injection. All three resulting mouse lines had the *neomycin* resistance gene in the intron next to the exon which was either wild-type or carried the R233Q or K236Q mutation. To ensure that all analysis was performed on precisely matched controls, all mutant mice were maintained as double heterozygotes of wild-type/mutant knockin on a mixed SV129/Black 6 background, and all analyses were performed on littermates derived from matings between double heterozygotes. Mice of various

genotypes and wild-type control mice were studied by immunoblotting for synaptic proteins as described²⁷.

Electrophysiological analyses of hippocampal neurons

Individual cultures of hippocampal neurons from all mice in a litter from double-heterozygous mutant/wild-type knockin mice were prepared at P0 or P1 on microislands of glia cells (pre-plated in 10% fetal bovine serum) under conditions favouring formation of autapses, and used for experiments after 10–20 d in culture^{30–32}. Before seeding neurons in a density of 500 per cm², the medium was exchanged to neurobasal medium A (Gibco) with supplement B27 (Gibco). Only dots containing single neurons were used. Only excitatory EPSCs were analysed since too few measurements were obtained from inhibitory GABAergic. The extracellular medium for the cultures contained (in mM): NaCl 140, KCl 2.4, HEPES 10, glucose 10, CaCl_2 4, MgCl_2 4; pH 7.3; 300 mOsm. For the measurements of mEPSCs, tetrodotoxin (200 nM) was added. NMDA EPSCs were measured in the presence of external 2.7 mM Ca^{2+} and 10 μM glycine but without Mg^{2+} . Synaptic transmission was recorded in the whole-cell configuration under voltage-clamp using 1–2 ms depolarizations from –75 mV to 0 mV to induce action potentials. Hypertonic sucrose solutions contained 0.5 M sucrose in addition to the regular external solution. Patch pipette solutions included (in mM): K-gluconate 125, NaCl 10, MgCl_2 4.6, ATP–Na₂ 4, creatine phosphate 15, phosphocreatine kinase (20 U ml^{–1}), EGTA 1; buffer pH 7.3; 300 mOsm. Chemicals were purchased from Sigma, except for MK-801 (Tocris).

Received 9 October; accepted 12 December 2000.

- Katz, B. *The Release of Neural Transmitter Substances* (Liverpool Univ. Press, Liverpool, 1969).
- Borst, J. G. & Sakmann, B. Calcium influx and transmitter release in a fast CNS synapse. *Nature* **383**, 431–434 (1996).
- Sabatini, B. L. & Regehr, W. G. Timing of synaptic transmission. *Annu. Rev. Physiol.* **61**, 521–542 (1999).
- Bennett, M. R. The concept of a calcium sensor in transmitter release. *Prog. Neurobiol.* **59**, 243–277 (1999).
- Burgoyne, R. D. & Morgan, A. Calcium sensors in regulated exocytosis. *Cell Calcium* **24**, 367–376 (1998).
- Kelly, R. B. Neural transmission. Synaptotagmin is just a calcium sensor. *Curr. Biol.* **5**, 257–259 (1995).
- von Gersdorff, H. & Matthews, G. Electrophysiology of synaptic vesicle cycling. *Annu. Rev. Physiol.* **61**, 725–752 (1999).
- Perin, M. S., Fried, V. A., Mignery, G. A., Jahn, R. & Südhof, T. C. Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. *Nature* **345**, 260–263 (1990).
- Geppert, M. *et al.* Synaptotagmin I: A major Ca^{2+} sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
- Sutton, R. B., Davletov, B. A., Berghuis, A. M., Südhof, T. C. & Sprang, S. R. Structure of the first C_2 -domain of synaptotagmin I: a novel Ca^{2+} /phospholipid-binding fold. *Cell* **80**, 929–938 (1995).
- Shao, X., Fernandez, L., Südhof, T. C. & Rizo, J. Solution structures of the Ca^{2+} -free and Ca^{2+} -bound C_2A -domain of synaptotagmin I: does Ca^{2+} induce a conformational change? *Biochemistry* **37**, 16106–16115 (1998).
- Ubach, J., Zhang, X., Shao, X., Südhof, T. C. & Rizo, J. Ca^{2+} binding to synaptotagmin: how many Ca^{2+} ions bind to the tip of a C_2 -domain? *EMBO J.* **17**, 3921–3930 (1998).
- Shao, X., Davletov, B. A., Sutton, R. B., Südhof, T. C. & Rizo, J. A bipartite Ca^{2+} -binding motif in C_2 -domains of synaptotagmin and protein kinase C. *Science* **273**, 248–251 (1996).
- Davletov, B. & Südhof, T. C. A single C_2 -domain from synaptotagmin I is sufficient for high affinity Ca^{2+} /phospholipid-binding. *J. Biol. Chem.* **268**, 26386–26390 (1993).
- Zhang, X., Rizo, J. & Südhof, T. C. Mechanism of phospholipid binding by the C_2A -domain of synaptotagmin I. *Biochemistry* **36**, 12395–12403 (1998).
- Brose, N., Petrenko, A. G., Südhof, T. C. & Jahn, R. Synaptotagmin: A Ca^{2+} sensor on the synaptic vesicle surface. *Science* **256**, 1021–1025 (1992).
- Mosier, M. & Epsand, R. M. Characterization of the calcium-binding site that regulates association of protein kinase C with phospholipid bilayers. *J. Biol. Chem.* **269**, 13798–13805 (1994).
- Verdaguer, N., Corbalan-Garcia, S., Ochoa, W. F., Fita, I. & Gomez-Fernandez, J. C. Ca^{2+} bridges the C_2 membrane-binding domain of protein kinase C α directly to phosphatidylserine. *EMBO J.* **18**, 6329–6338 (2000).
- Li, C. *et al.* Ca^{2+} -dependent and Ca^{2+} -independent activities of neural and nonneural synaptotagmins. *Nature* **375**, 594–599 (1995).
- Chapman, E. R., Hanson, P. I., An, S. & Jahn, R. Ca^{2+} regulates the interaction between synaptotagmin and syntaxin I. *J. Biol. Chem.* **270**, 23667–23771 (1995).
- Kee, Y., Scheller, R. H. Localization of synaptotagmin-binding domains on syntaxin. *J. Neurosci.* **16**, 1975–1981 (1996).
- Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* **17**, 957–967 (1996).
- Schneggenburger, R. & Neher, E. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature* **406**, 889–893 (2000).
- Bollmann, J. H., Sakmann, B. & Borst, J. G. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* **289**, 953–957 (2000).
- Garcia, R. A., Forde, C. E. & Godwin, H. A. Calcium triggers an intramolecular association of the C_2 -domains in synaptotagmin. *Proc. Natl Acad. Sci. USA* **97**, 5883–5888 (2000).
- Shao, X. *et al.* Synaptotagmin–syntaxin interaction: the C_2 -domain as a Ca^{2+} -dependent electrostatic switch. *Neuron* **18**, 133–142 (1997).
- Rosahl, T. W. *et al.* Essential functions of synapsins I and II in synaptic vesicle regulation. *Nature* **375**, 488–493 (1995).
- Perin, M. S., Brose, N., Jahn, R. & Südhof, T. C. Domain structure of synaptotagmin. *J. Biol. Chem.* **266**, 623–629 (1991).
- Tugal, H. B., van Leeuwen, F., Apps, D. K., Haywood, J. & Phillips, J. H. Glycosylation and transmembrane topography of bovine chromaffin granule p65. *Biochem. J.* **279**, 699–703 (1991).
- Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).

31. Gomperts, S. N., Rao, A., Craig, A. M., Malenka, R. C. & Nicoll, R. A. Postsynaptically silent synapses in single neuron cultures. *Neuron* **21**, 443–451 (1998).
32. Augustin, I., Rosenmund, C., Südhof, T. C. & Brose, N. Munc-13 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature* **400**, 457–461 (1999).
33. Zucker, R. S. Short-term synaptic plasticity. *Annu. Rev. Neurosci.* **12**, 13–31 (1989).
34. Thomson, A. M. Facilitation, augmentation and potentiation at central synapses. *Trends Neurosci.* **23**, 305–312 (2000).
35. Rosenmund, C., Clements, J. D. & Westbrook, G. L. Nonuniform probability of glutamate release at a hippocampal synapse. *Science* **262**, 754–757 (1993).
36. Hessler, N. A., Shirke, A. M. & Malinow, R. The probability of transmitter release at a mammalian central synapse. *Nature* **366**, 569–572 (1993).
37. Rosenmund, C. & Stevens, C. F. Definition of the readily releasable pool of vesicles at hippocampal synapses. *Neuron* **16**, 1197–1207 (1996).
38. Fatt, P. & Katz, B. Spontaneous subthreshold activity at motor nerve endings. *J. Physiol. (Lond.)* **117**, 109–128 (1952).
39. Mintz, I. M., Sabatini, B. L. & Regehr, W. G. Calcium control of transmitter release at a cerebellar synapse. *Neuron* **15**, 675–688 (1995).
40. Catterall, W. A. Interactions of presynaptic Ca^{2+} -channels and SNARE proteins in neurotransmitter release. *Ann. N. Y. Acad. Sci.* **868**, 144–159 (1999).
41. Seagar, M. *et al.* Interactions between proteins implicated in exocytosis and voltage-gated calcium channels. *Phil. Trans. R. Soc. Lond. B* **354**, 289–297 (1999).
42. Uellner, R. *et al.* Perforin is activated by a proteolytic cleavage during biosynthesis which reveals a phospholipid-binding C_2 -domain. *EMBO J.* **16**, 7287–7296 (1997).
43. Edwards, A. S. & Newton, A. C. Regulation of protein kinase C β II by its C_2 -domain. *Biochemistry* **36**, 15615–15623 (1997).
44. Nalefski, E. A. & Falke, J. J. Location of the membrane-docking face on the Ca^{2+} -activated C_2 -domain of cytosolic phospholipase A2. *Biochem.* **37**, 17642–17650 (1998).
45. Williams, R. L. & Katan, M. Structural views of phosphoinositide-specific phospholipase C: signalling the way ahead. *Structure* **4**, 1387–1394 (1996).
46. Blasi, J. *et al.* Botulinum neurotoxin C1 blocks neurotransmitter release by means of cleaving HPC-1/syntaxin. *EMBO J.* **12**, 4821–4828 (1993).
47. Sugita, S., Hata, Y. & Südhof, T. C. Distinct Ca^{2+} -dependent properties of the first and second C_2 -domains of synaptotagmin I. *J. Biol. Chem.* **271**, 1262–1265 (1996).
48. Chapman, E. R., An, S., Edwardson, J. M. & Jahn, R. A novel function for the second C_2 -domain of synaptotagmin. Ca^{2+} -triggered dimerization. *J. Biol. Chem.* **271**, 5844–5849 (1996).
49. Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. Backbone ^1H and ^{15}N resonance assignments of N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* **4**, 845–858 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank I. Herfort, I. Leznicki and A. Roth for technical assistance; H. Riedesel, J. Krause, S. Röcklin and the ARC in Dallas for help with mouse husbandry; and E. Neher, G. Alvarez de Toledo, J. López-Barneo and R. Jahn for advice. This study was supported by grants from the NIH to J.R., a grant from the Deutsche Forschungsgemeinschaft to C.R., Heisenberg Fellowships from the Deutsche Forschungsgemeinschaft to N.B. and C.R., a grant from the Perot Family Foundation to T.C.S., and postdoctoral fellowships from the Spanish Ministry of Education and the Fulbright Commission to R.F.C., and from the Deutsche Forschungsgemeinschaft to S.H.G.

Correspondence and requests for materials should be addressed to T.C.S. (e-mail: Thomas.Sudhof@UTSouthwestern.edu).

Involvement of chemokine receptors in breast cancer metastasis

Anja Müller^{*†‡}, Bernhard Homey^{*†‡}, Hortensia Soto^{*}, Nianfeng Ge^{*}, Daniel Catron^{*}, Matthew E. Buchanan^{*}, Terri McClanahan^{*}, Erin Murphy^{*}, Wei Yuan^{*}, Stephan N. Wagner[§], Jose Luis Barrera^{||}, Alejandro Mohar[¶], Emma Verástegui^{||} & Albert Zlotnik^{*}

^{*} Department of Immunology, DNAX Research Institute, 901 California Avenue, Palo Alto, California 94304, USA

[†] Departments of Radiation Oncology and Dermatology, Heinrich-Heine University, Moorenstrasse 5, D-40225 Düsseldorf, Germany

[§] Department of Dermatology, University of Essen, Hufelandstrasse 55, D-45147 Essen, Germany

^{||} Instituto Nacional de Cancerología, Av. San Fernando 22, Tlalpan 14000 D.F., México

[¶] Instituto de Investigaciones Biomédicas, Ciudad Universitaria, 04510 D.F., México

[‡] These authors contributed equally to this work

Breast cancer is characterized by a distinct metastatic pattern involving the regional lymph nodes, bone marrow, lung and liver. Tumour cell migration and metastasis share many similarities with leukocyte trafficking, which is critically regulated by chemokines and their receptors. Here we report that the chemokine receptors CXCR4 and CCR7 are highly expressed in human breast cancer cells, malignant breast tumours and metastases. Their respective ligands CXCL12/SDF-1 α and CCL21/6CKine exhibit peak levels of expression in organs representing the first destinations of breast cancer metastasis. In breast cancer cells, signalling through CXCR4 or CCR7 mediates actin polymerization and pseudopodia formation, and subsequently induces chemotactic and invasive responses. *In vivo*, neutralizing the interactions of CXCL12/CXCR4 significantly impairs metastasis of breast cancer cells to regional lymph nodes and lung. Malignant melanoma, which has a similar metastatic pattern as breast cancer but also a high incidence of skin metastases, shows high expression levels of CCR10 in addition to CXCR4 and CCR7. Our findings indicate that chemokines and their receptors have a critical role in determining the metastatic destination of tumour cells.

Metastasis is the result of several sequential steps and represents a highly organized, non-random and organ-selective process¹. Although a number of molecules have been implicated in the metastasis of breast cancer, the precise mechanisms determining the directional migration and invasion of tumour cells into specific organs remain to be established^{1–3}.

Chemokines are a superfamily of small, cytokine-like proteins that induce, through their interaction with G-protein-coupled receptors, cytoskeletal rearrangement, firm adhesion to endothelial cells and directional migration^{4–6}. These secreted proteins act in a coordinated fashion with cell-surface proteins, including integrins, to direct the specific homing of various subsets of haematopoietic cells to specific anatomical sites^{4–10}.

We therefore thought that tumour cells may use chemokine-mediated mechanisms such as those regulating leukocyte trafficking during the process of metastasis. Here we report that tumour cells express a distinct, non-random pattern of functionally active chemokine receptors. Signalling through CXCR4 or CCR7 mediates actin polymerization and pseudopodia formation in breast cancer cells, and induces chemotactic and invasive responses. In addition, we find that organs representing the main sites of breast cancer metastasis are the most abundant sources of ligands for these tumour-associated receptors. *In vivo*, neutralizing the interactions of CXCL12/CXCR4 leads to a significant inhibition of lymph-node and lung metastasis.

Expression of chemokine receptors in tumour cells

To determine whether chemokine/chemokine receptor interactions are involved in the metastatic process, we performed a comprehensive, quantitative analysis of the expression of all known chemokine receptors (CCR1–CCR10, CXCR1–CXCR5, XCR1 and CX₃CR1) in seven human breast cancer cell lines, and compared expression in normal primary mammary epithelial cells. Absolute messenger RNA levels determined using real-time quantitative polymerase chain reaction (PCR) showed that breast cancer cells express chemokine receptors in a defined rather than a random manner (Fig. 1a).

Three different patterns of receptor expression were observed. First, the chemokine receptor was expressed in normal mammary epithelial cells, but uniformly downregulated by breast cancer cells (for example, CXCR2). Second, both normal mammary epithelial cells and malignant cells expressed significant levels of receptor mRNA; this pattern was detected for CCR7 and CCR8, with CCR7 showing the most consistent upregulation in breast cancer cells. Third, receptor expression was undetectable in normal mammary epithelial cells but markedly upregulated in breast cancer cells (for example, CXCR4).

Flow cytometric analyses confirmed strong cell-surface expression of CXCR4 on breast cancer cell lines (data not shown), and on primary breast cancer cells (82.33–97.98% CXCR4-positive malignant cells) recovered from patients with malignant pleural effusion ($n = 5$) (see Supplementary Information).

We quantitatively measured CCR7 and CXCR4 mRNA expression in primary tumours of human invasive lobular or ductal breast carcinoma ($n = 12$), and in normal mammary gland ($n = 5$) and its cellular constituents (Fig. 2a,b). According to the TNM classification¹¹, six tumours were classified as T₂, four as T₃ and two as T₄, with evidence of axillary nodal metastasis in five cases. Figure 2a and b shows that mRNA of both CCR7 and CXCR4 was significantly upregulated in primary breast tumours compared with normal mammary gland tissue ($P < 0.05$, $P < 0.005$, respectively). Normal primary mammary epithelial and stromal cells derived from three different donors did not express detectable levels of CXCR4 mRNA. There was no significant correlation between the TNM status of the 12 patients studied and absolute levels of CCR7 and CXCR4 mRNA expression in primary tumours.

We confirmed *in vivo* protein expression of CXCR4 by immunohistochemistry (Fig. 2c–n). Stained tissue sections from normal mammary gland ($n = 4$) (Fig. 2c–e) or invasive ductal carcinoma ($n = 8$) (Fig. 2f–k) indicated that breast cancer cells express high levels of CXCR4, but normal mammary ductal cells do not express this receptor. In addition to primary tumours, axillary lymph-node metastases ($n = 5$) (Fig. 2l–n) and distant metastases of lung and liver ($n = 8$) (data not shown) exhibited strong CXCR4 expression

in tumour cells.

Although CXCL12/SDF-1 mRNA has been reported to be expressed in many different tissues¹², quantitative analysis of CXCL12 expression in many human organs revealed that CXCL12 mRNA is expressed preferentially in lymph nodes, lung, liver and bone marrow, and shows markedly lower expression in small intestine, kidney, skin, brain and skeletal muscle (Fig. 2o). Organs exhibiting the highest CXCL12 expression represent the most common sites of metastasis in breast cancer¹³.

In addition, both CCL21/6CKine and CCL19/MIP-3 β had

showed most abundant expression in lymph nodes; however, CCL21 was expressed at higher levels, supporting its critical role in the homing of cells into lymph nodes^{14,15} (Fig. 2p; and see Supplementary Information). Comparing CXCL12 and CCL21 expression in primary breast tumours and breast cancer cell lines with that in normal human organs showed that tumour cells express low or undetectable amounts of mRNA of these particular chemokines (see Supplementary Information).

To determine whether the distinct expression of chemokine receptors is unique to breast cancer, we also analysed malignant melanoma, which, like breast cancer, preferentially develops lymph-node, lung, liver and bone-marrow metastases, but also has a high frequency of skin metastases. In addition to CXCR4 and CCR7, malignant melanoma cells expressed high levels of CCR10 mRNA as compared with normal primary melanocytes (Fig. 1b). Notably, the CCR10 ligand CCL27/CTACK represents a skin-specific homeostatic chemokine (Fig. 2q) that has been associated with the homing of memory T cells into the skin^{7,8}.

Breast cancer cells express active CXCR4 and CCR7

In vitro, chemokine ligand–receptor interactions trigger intracellular actin polymerization in leukocytes, a process that is prerequisite for cell motility and migration^{16,17}. Consistent with findings in leukocytes¹⁶, CXCL12 (100 nM) and CCL21 (100 nM) induced, respectively, a transient 2.2- and 1.6-fold increase in intracellular filamentous actin (F-actin) in human breast cancer cells within 20 s (Fig. 3a). Conversely, the chemokine CX₃CL1/fractalkine, whose receptor CX₃CR1/V28 was not detected on breast cancer cells (Fig. 1a), did not induce actin polymerization (Fig. 3a).

In tumour cells, high levels of actin polymerization are required for the formation of pseudopodia, which in turn are needed for the invasion of malignant cells into tissues and for efficient metastases formation¹⁸. Confocal laser scan microscopy of breast cancer cells stimulated in suspension with either CXCL12 or CCL21 revealed intense F-actin staining in the periphery of the cells and a redistribution of F-actin towards a leading edge (Fig. 3b–e). In adherent breast cancer cells, distinct pseudopodia formation was observed after 20 min of stimulation with either CCL21 or CXCL12 (Fig. 3f, g).

In agreement with these findings, both CXCL12 and CCL21 induced directional migration of breast cancer cells (Fig. 4a–d) and directional invasion through a reconstituted basement membrane (Fig. 4e, f) in a dose-dependent manner. Optimal migratory/invasive responses to CXCL12 (Fig. 4a, c, e) or CCL21 (Fig. 4b, d, f) were observed at concentrations of 100 nM, or 100 and 200 nM, respectively, reminiscent of observations made with leukocytes^{16,19,20}. Compared with breast cancer cells of well-characterized cell lines (MDA-MB-231, MDA-MB-361), primary tumour cells derived from a patient with malignant pleural effusion exhibited significant chemotactic responses to both CXCL12 and CCL21 (Fig. 4c, d). CXCL12- and CCL21-mediated chemotaxis and invasion could be blocked by neutralizing anti-CXCR4 or anti-CCL21 antibodies, respectively, confirming the specificity of the chemotactic response induced by these chemokines (data not shown).

Migratory response to organ-derived proteins

To determine the biological relevance of CXCL12-mediated chemotaxis in the context of all chemotactic factors present in the extracellular matrix of a particular organ, protein extracts of normal human lung, liver, skin and muscle, and conditioned media from human primary bone-marrow and lymph-node stromal cells, were tested for their chemotactic activity on breast cancer cells (Fig. 5a–f). Protein extracts of lung and liver induced chemotactic responses in breast cancer cells, indicating that extracts derived from these organs contain factors with chemotactic properties (data not shown). Blocking the CXCL12/CXCR4 interactions with neutralizing anti-CXCR4 (Fig. 5a, b) or anti-CXCL12 antibodies (see Supplementary Information) significantly impaired

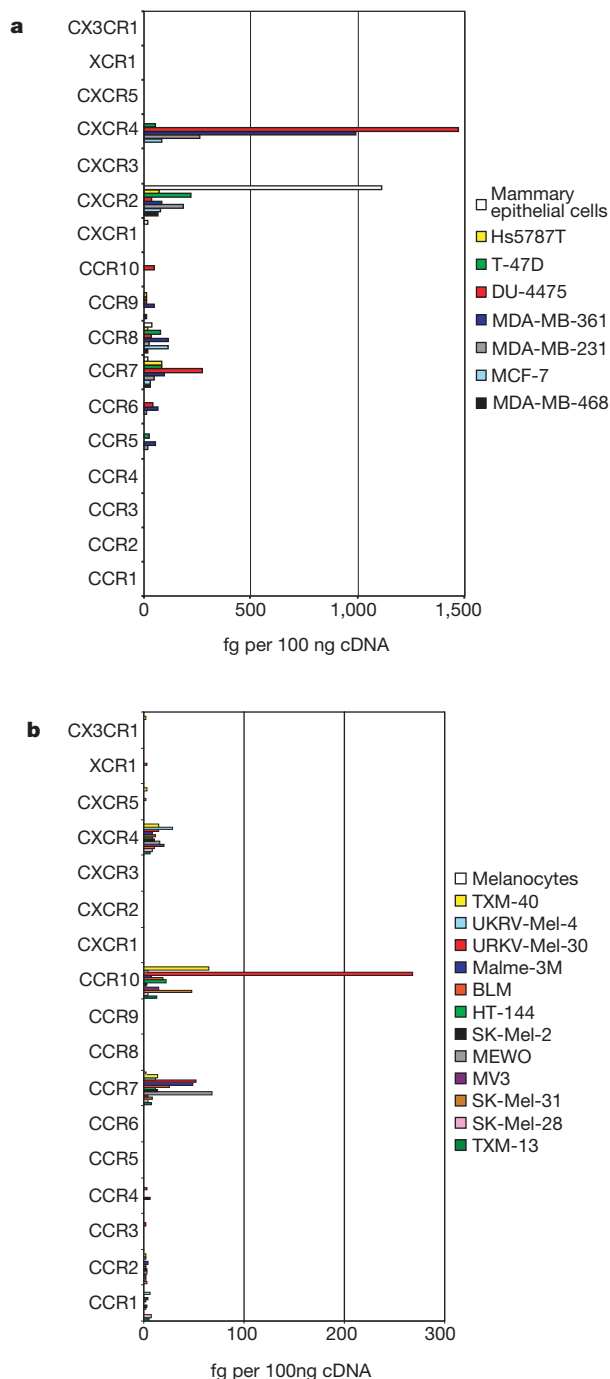


Figure 1 Expression of chemokine receptors in human tumour cells. Quantitative RT–PCR analyses of all known chemokine receptors in seven human breast cancer cell lines compared with normal primary mammary epithelial cells (**a**), and in 12 human malignant melanoma cell lines compared with normal primary melanocytes (**b**). Values are expressed as femtograms of target gene in 100 ng of total cDNA.

these migratory responses by 63–76% and 60–62%, respectively, indicating that CXCL12 is one of the main chemotactic factors for breast cancer cells in the extracellular matrix of these organs.

Furthermore, migration of breast cancer cells in response to conditioned medium from human bone-marrow (Fig. 5c) or lymph-node (Fig. 5d) stromal cells was significantly reduced (53–61% and 51–54%, respectively) after CXCR4 blocking. In contrast, protein extracts from organs that represent rare targets of breast cancer metastasis, such as skin (Fig. 5e) or muscle (Fig. 5f), exhibited weak overall chemoattractive properties for breast cancer cells, compared with protein extracts from lung and liver. Notably, these migratory responses were not affected by CXCR4 neutralization (Fig. 5e, f).

CXCR4-neutralization inhibits metastasis *in vivo*

Having established that breast cancer cells express functionally active CXCR4, we evaluated the contribution of CXCL12/CXCR4

interactions on metastasis *in vivo* by using experimental and spontaneous metastasis models of breast cancer. The CXCR4-positive, human breast carcinoma cell line MDA-MB-231 was injected either intravenously (i.v.) into the tail vein or orthotopically into the mammary fat pad of severe combined immunodeficient (SCID) mice.

In vitro, chemotaxis and invasion assays confirmed that MDA-MB-231 cells respond to murine CXCL12 (data not shown). Real-time quantitative PCR analyses showed that there was significant expression of human CXCR4 mRNA in both primary tumours and metastases-infiltrated lungs of SCID mice that had been injected with MDA-MB-231 cells (Fig. 6a). Immunohistochemical staining for human CXCR4 depicted strong protein expression by tumour cells of primary tumours (Fig. 6c, d) and lung metastases (Fig. 6e, f), indicating that MDA-MB-231 cells maintained high levels of CXCR4 expression *in vivo*. In agreement with CXCL12 mRNA expression in normal human organs (Fig. 2o), quantitative analysis

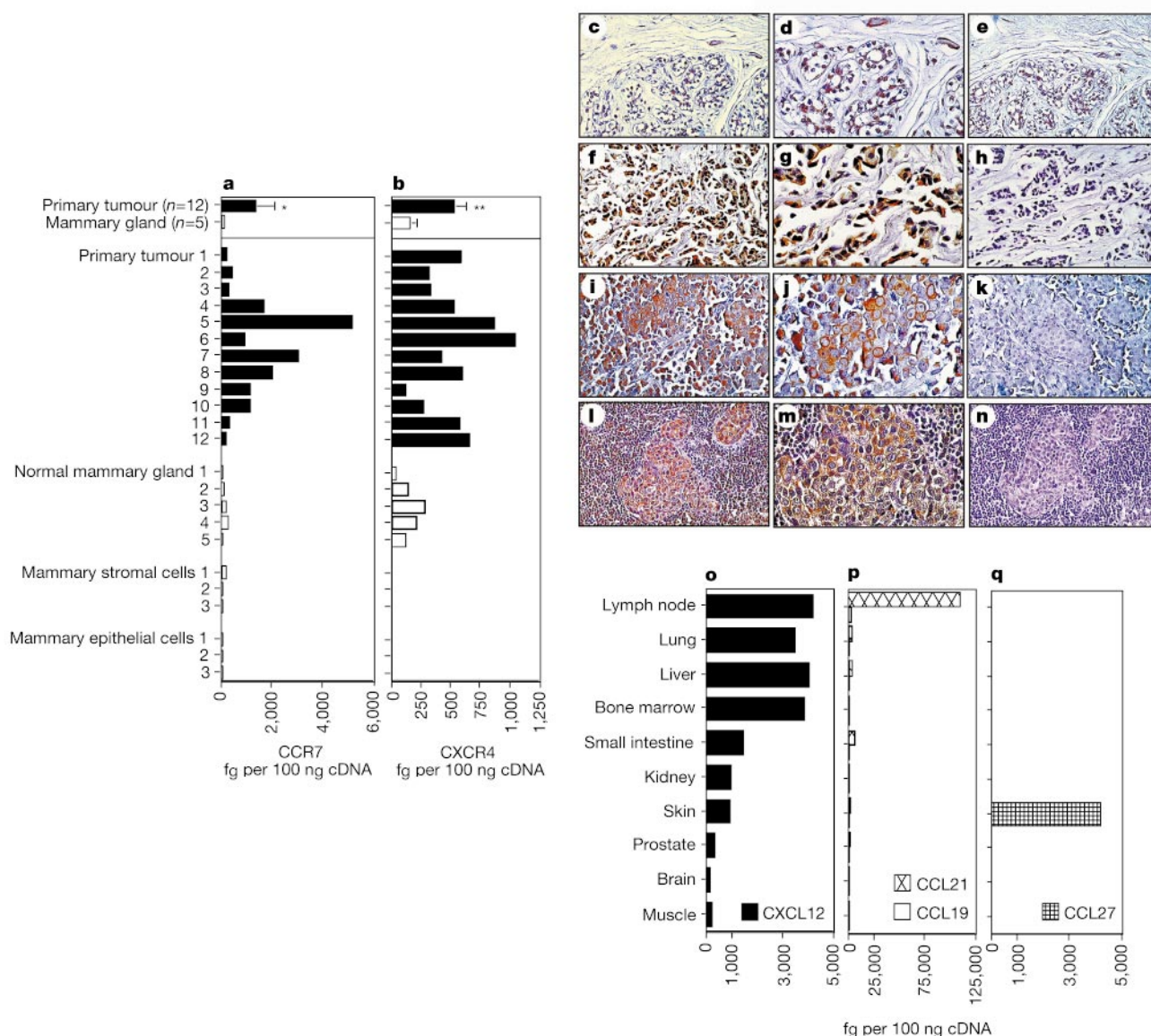


Figure 2 Expression of CCR7 and CXCR4 in breast cancer. **a, b**, Quantitative RT-PCR analyses of CCR7 (**a**) and CXCR4 (**b**) mRNA expression in primary breast carcinoma ($n = 12$), normal mammary gland tissue ($n = 5$), primary mammary epithelial ($n = 3$) and stromal ($n = 3$) cells. Mean $\pm$ s.e.m. (Student's *t*-test; * $P < 0.05$, ** $P < 0.005$). **c–n**, Immunohistochemical evaluation of CXCR4 expression. **c–e**, Section of normal mammary gland tissue. **f–h, i–k**, Invasive ductal carcinoma of two different patients.

l–n, Axillary lymph-node metastasis. Original magnification, $\times 200$ (**c, f, i, l**, anti-CXCR4; **e, h, k, n**, isotype); $\times 400$ (**d, g, j, m**, anti-CXCR4). Quantitative RT-PCR analyses of CXCL12/SDF-1 (**o**), CCL21/6CKine and CCL19/MIP-3 β (**p**) or CCL27/CTACK (**q**) mRNA expression in various normal human organs. Values are expressed as femtograms of target gene in 100 ng of total cDNA.

of murine CXCL12 mRNA in normal organs of SCID mice showed peak levels of expression in mouse lung, lymph nodes, bone marrow and liver (data not shown).

Twenty-eight days after i.v. injection of MDA-MB-231 cells and the twice weekly treatment with either neutralizing anti-human CXCR4 monoclonal antibody or isotype control, a significant

decrease in lung metastasis was observed in anti-CXCR4-treated mice (relative suppression, 61–68%; $P < 0.001$) (Fig. 7a–c, e). Furthermore, spontaneous metastasis to the lung 44 days after orthotopic injection of MDA-MB-231 cells was impaired significantly by treatment with a neutralizing anti-CXCR4 monoclonal antibody (relative suppression, 73–82%; $P < 0.001$) (Fig. 7d, f).

In addition to the histological evaluation of lung infiltrates, the individual tumour burden per lung was measured by quantitative PCR using primers specific for the human housekeeping gene *HPRT*. Expression of human *HPRT* was detectable in cultured human MDA-MB-231 breast carcinoma cells, primary tumours and tumour-infiltrated lungs of untreated SCID mice, but undetectable in normal mouse lungs (Fig. 6b). The human *HPRT* mRNA content correlated directly with the tumour load of each sample. Subsequent quantitative PCR analyses of human *HPRT* in lungs of either isotype- or anti-CXCR4-treated mice indicated a significant reduction in lung-infiltrating tumour cells after anti-CXCR4 treatment in both experimental (i.v.) and spontaneous metastasis models (relative suppression, 61–73%; $P < 0.05$) (Fig. 7e, f). Furthermore, levels of human CXCR4 mRNA in the lungs of CXCR4-treated mice showed a relative suppression, ranging from 67 to 89%, as compared with isotype-treated mice (Fig. 7g, h). Thus, quantitative PCR analyses confirmed the histological quantification.

Clinical evaluation of draining lymph nodes 44 days after orthotopic injection of MDA-MB-231 cells showed that anti-CXCR4 treatment inhibited metastasis to inguinal and axillary lymph nodes

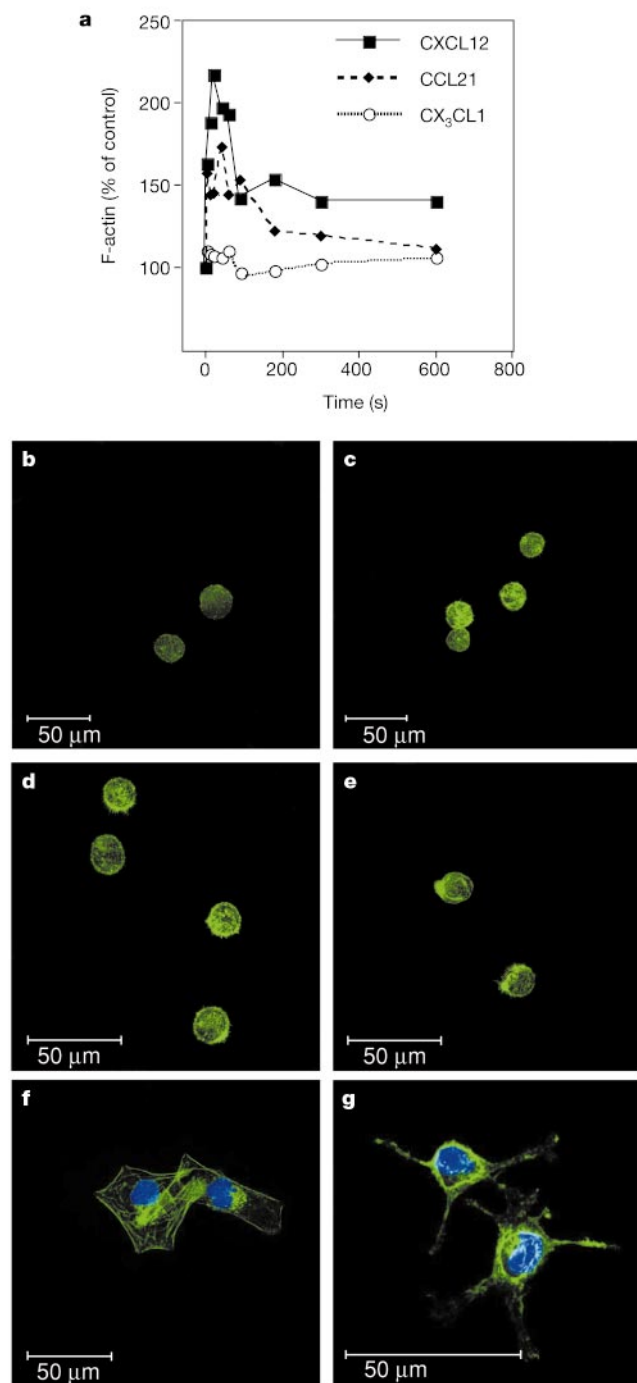


Figure 3 F-actin polymerization in human breast carcinoma cells. **a**, Intracellular F-actin content measured by flow cytometry in MDA-MB-231 cells after stimulation with CXCL12/SDF-1 α (100 nM), CCL21/6Ckine (100 nM) or CX₃CL1/fraktalkine (100 nM). Data points are plotted relative to the mean fluorescence before the addition of chemoattractant. **b–e**, Confocal microscopy series of breast carcinoma cells after CXCL12 (100 nM) stimulation in suspension. MDA-MB-231 cells unstimulated (**b**) or stimulated for 20 s (**c**), 3 min (**d**) or 20 min (**e**). **f, g**, Adherent MDA-MB-231 cells unstimulated (**f**) or stimulated (**g**) with CXCL12 (100 nM) for 20 min. Scale bar, 50 μ m.

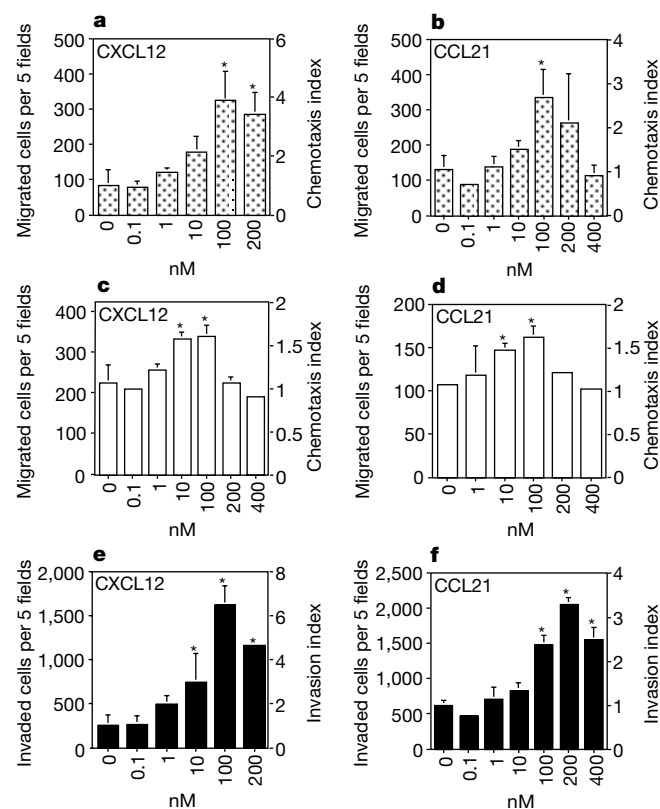


Figure 4 Chemokine-mediated migration and invasion of breast carcinoma cells. Chemotactic response of MDA-MB-231 cells to different concentrations of CXCL12/SDF-1 α (**a**) or CCL21/6Ckine (**b**). Directional migration of primary breast carcinoma cells towards CXCL12 (**c**) or CCL21 (**d**) gradients. Effects of various concentrations of CXCL12 (**e**) or CCL21 (**f**) on the invasion of MDA-MB-361 cells. Results are expressed as the mean number of migrating cells per well. Chemotaxis indices were calculated as ratio of cells migrated toward a chemokine gradient to cells migrated in the negative control. Mean $\pm$ s.d. (Student's *t*-test; * $P < 0.005$).

(Table 1). All isotype-treated mice developed ipsi- and contralateral inguinal, and axillary lymph-node metastases at the site of injection (ipsilateral) with diameters exceeding 3 mm. In contrast, only 38% of anti-CXCR4-treated mice presented inguinal lymph-node metastases at the site of injection (diameter < 3 mm) (Table 1).

Discussion

It has been proposed that molecules regulating the metastatic dissemination of tumour cells to specific anatomical sites need to fulfil the following criteria^{1,2}. First, they have to be constitutively expressed at principal sites of metastasis. Second, adhesion of target cells to the endothelium and transendothelial migration need to be promoted. Third, these molecules must be capable of mediating the invasion of cells into tissues that provide supportive microenvironments. Last, this process requires the expression of a distinct receptor repertoire by the target cells, depending on their metastatic profile.

Given their well-established roles in leukocyte trafficking and

homeostasis, chemokines are perfectly positioned to fulfil these criteria^{4,5,7–10,21,22}. We have shown that, out of all known chemokine receptors, breast cancer cells specifically express functionally active CXCR4 and CCR7, which trigger actin polymerization, pseudopodia formation, and the directional migration and invasion of breast cancer cells. Together with the distinct tissue distribution of their ligands CXCL12 and CCL21 in organs representing the main sites of breast cancer metastasis and the significant inhibition of breast cancer metastasis observed *in vivo*, this finding indicates that chemokine ligand–receptor interactions may be crucial for the metastasis of breast cancer.

Although chemokines have been implicated in tumour cell growth^{2,23}, angiogenesis²⁴ and the host immune response against malignant cells^{2,25}, our study provides evidence that chemokines may promote metastasis by acting directly on tumour cell migration and invasion. Both CXCR4 and CCR7 are critical for cell trafficking and tissue homeostasis^{4,9,10,26–29}. CXCL12 is the only known ligand for CXCR4 (refs 26–29). This chemokine ligand–receptor pair

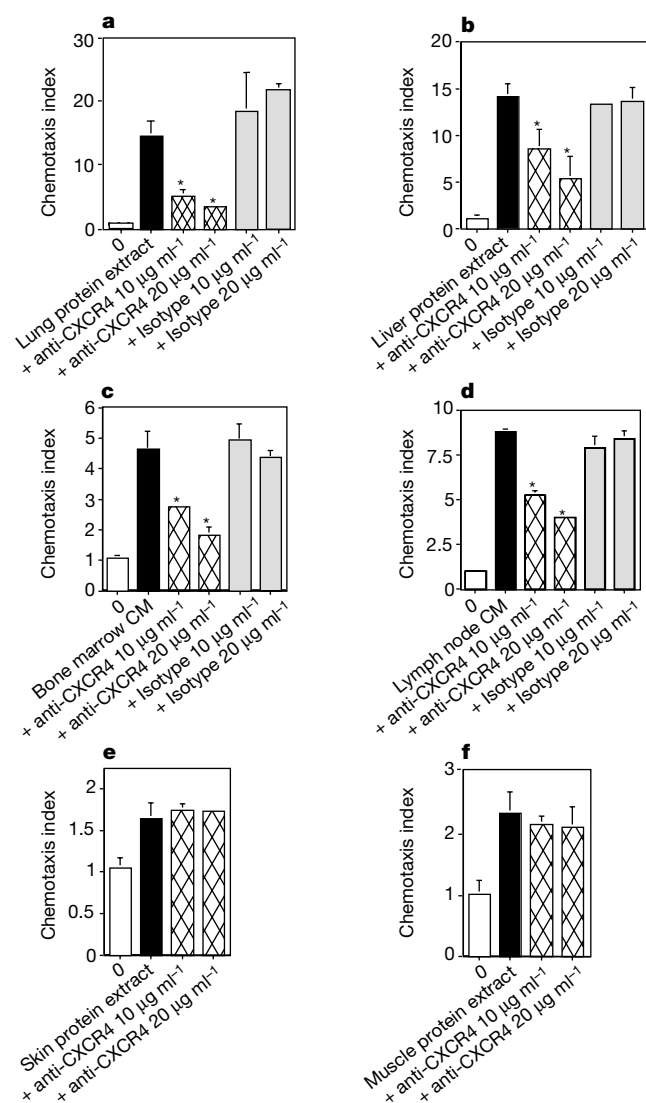


Figure 5 Migration of breast carcinoma cells in response to organ-derived proteins. Chemotaxis of MDA-MB-231 breast carcinoma cells in response to protein extracts of human lung (a) or liver (b), conditioned medium (CM) from human primary bone marrow (c) or lymph-node stromal cells (d) and protein extracts from human skin (e) or muscle (f) in the presence of a neutralizing anti-CXCR4 monoclonal antibody or isotype control. Chemotaxis indices were calculated as ratio of cells migrated toward a protein gradient to cells migrated in the negative control. Mean $\pm$ s.d. (Student's *t*-test; **P* < 0.001).

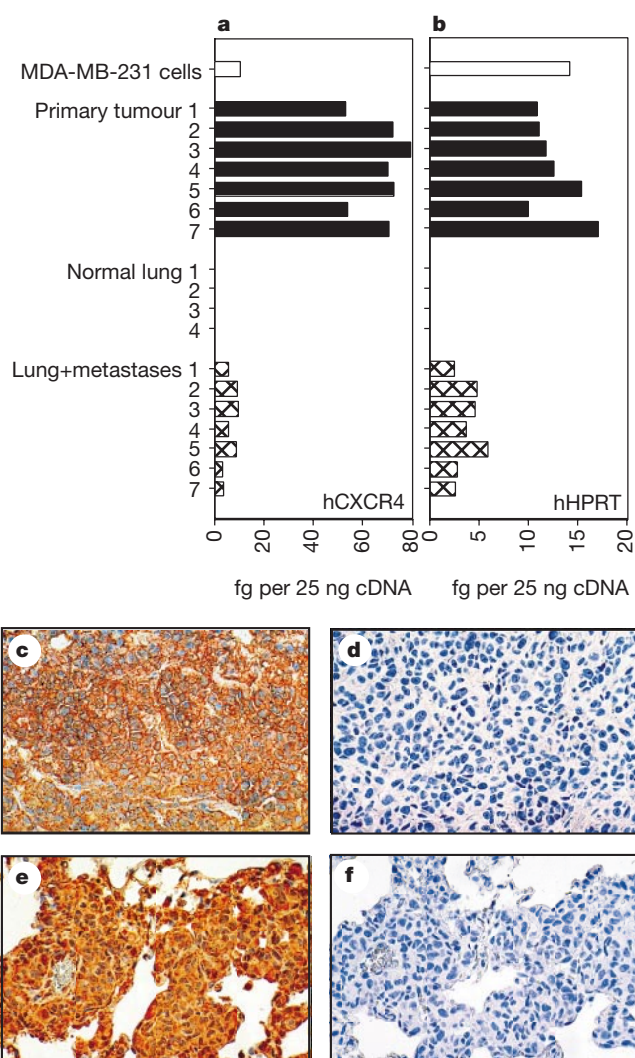


Figure 6 Expression of human CXCR4 and human HPRT in the MDA-MB-231 breast cancer metastasis model. Quantitative real-time RT-PCR (TaqMan) analyses of hCXCR4 (a) and hHPRT (b) mRNA expression in primary tumours (*n* = 7), metastases-infiltrated lungs (*n* = 7), and normal lungs of SCID mice. Values are expressed as femtograms of target gene in 25 ng of total cDNA. Immunohistochemistry of human CXCR4 in primary tumours (c, d) and lung metastases (e, f). Representative staining in tissue samples from one out of six mice. Original magnification, $\times 400$ (c, e, anti-CXCR4; d, f, isotype).

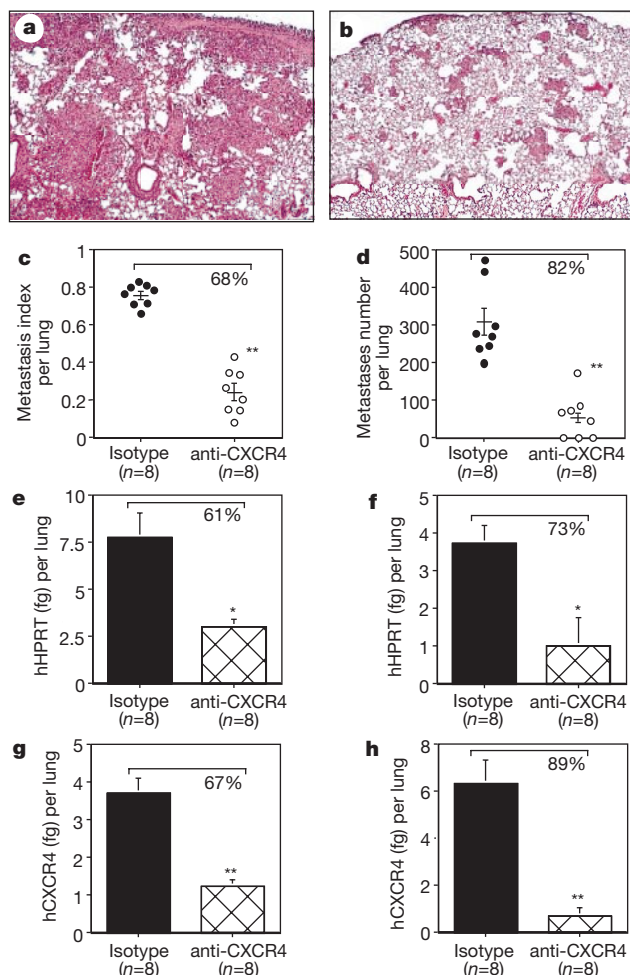


Figure 7 Effect of CXCR4-neutralization on metastasis *in vivo*. **a, b**, Haematoxylin and eosin staining of lungs from SCID mice injected with human MDA-MB-231 breast carcinoma cells and treated with either isotype control (**a**) or anti-hCXCR4 monoclonal antibody (**b**). Original magnification, $\times 100$. **c**, Lung colony formation after i.v. injection of MDA-MB-231 cells in isotype- or anti-CXCR4-treated mice. Representative data of one out of two experiments. **d**, Spontaneous lung metastasis after orthotopic injection of MDA-MB-231 cells. **e, f**, Metastases quantification by quantitative RT-PCR analyses in contralateral lungs. Expression of hHPRT mRNA after i.v. (**e**) or orthotopic (**f**) injection. Expression of hCXCR4 mRNA after i.v. (**g**) or orthotopic (**h**) injection. Values are expressed as femtograms of target gene in 25 ng of total cDNA. Mean $\pm$ s.e.m. (Student's *t*-test; $*P < 0.05$, $**P < 0.001$). Percentages indicate relative suppression compared with isotype.

exhibits unparalleled chemotactic efficacy for leukocytes *in vitro* and is a highly potent chemoattractant *in vivo*^{16,19,26–29}. Both CXCR4- and CXCL12-deficient mice display perinatal lethality owing to profound defects in embryonic development of the haematopoietic, cardiovascular and nervous systems^{26–29}. These phenotypic changes are mediated by the disrupted migration of embryonic progenitor cells into appropriate microenvironments, suggesting that CXCL12/CXCR4 interactions are vital for the migration of haematopoietic and non-haematopoietic cells *in vivo*.

Furthermore, *in vivo* studies using neutralizing antibodies to CXCR4 implicate this receptor in the homing and repopulation of human stem cells into the bone marrow of SCID mice⁹. As expression of CXCR4 by breast cancer cells is related to efficient, CXCL12-induced migration and invasion of these cells both *in vitro* and *in vivo*, it is conceivable that CXCL12/CXCR4 interactions may contribute to bone-marrow infiltration by breast cancer cells. The

Table 1 Frequency of lymph-node metastases

Mice*	Ipsilateral inguinal LN	Contralateral inguinal LN	Axillary LN
Isotype (n = 8)	8/8 (>3 mm)	8/8 (>3 mm)	8/8 (>3 mm)
Anti-CXCR4 (n = 8)	3/8 (<3 mm)	0/8	0/8

*MDA-MB-231 breast carcinoma cells (10^6) were injected into the inguinal mammary fat pad of SCID mice, which were treated with anti-CXCR4 monoclonal antibody or isotype control. Regional lymph-node (LN) metastasis was evaluated on day 44.

preferential expression of CXCL12 in lung, liver and lymph node also suggests a role for this chemokine in the tropism of breast cancer cells for these anatomical sites.

Data point to the crucial role of CCL21 and its receptor CCR7 in the homing of lymphocytes into secondary lymphoid organs. A natural mutation in mice, designated *plt* (for paucity of lymph-node T cells) that results in the loss of one of the forms of mCCL21 (refs 14, 15, 30), and targeted disruption of the CCR7 gene¹⁰ cause impaired homing of naive T cells to secondary lymphoid organs. Thus, the abundant expression of the homeostatic chemokine CCL21 in lymph nodes makes it a likely candidate to attract CCR7-positive tumour cells. In fact, CXCL12 and CCL21 may have synergistic effects, as both chemokines are highly expressed in lymph nodes and mediate chemotactic and invasive responses of breast cancer cells *in vitro*.

Our findings are probably not unique to breast cancer. Other tumour entities of haematopoietic and non-haematopoietic origin, including acute myeloid and lymphoblastic leukaemia³¹, chronic lymphocytic leukaemia^{17,32}, non-Hodgkin B-cell lymphoma³³ and pancreatic cancer³⁴, express functionally active chemokine receptors that mediate tumour cell migration *in vitro*. Our results in breast cancer and malignant melanoma suggest that malignant cells, in general, express distinct and non-random patterns of chemokine receptors. Furthermore, the association of CCR10 expression by malignant melanoma cells with the skin-specific expression of its ligand CCL27/CTACK^{7,8} and the high incidence of skin metastases in this malignant disease support the involvement of chemokine receptors in metastasis.

Currently, intense efforts are underway to identify small-molecule antagonists for many chemokine receptors³⁵. We propose that small molecule antagonists of chemokine receptors, such as CXCR4, may be useful to interfere with tumour progression and metastasis in tumour patients. □

Methods

Cell lines

The following human breast cancer and melanoma cell lines were obtained from the ATCC: DU-4475, MCF-7, T-47D, Hs578T, MDA-MB-468, MDA-MB-361, MDA-MB-231, SK-Mel-2, SK-Mel-28, SK-Mel-31, HT-144 and Malm-3M. MV3 melanoma cells were a gift from D. J. Ruiter, Nijmegen; TXM-14 and TXM-30 cells were obtained from I. J. Fidler, Houston; UKRV-Mel-4 and UKRV-Mel-30 melanoma cells were a gift from D. Schadendorf, Mannheim. Human primary mammary epithelial and stromal cells, and human primary melanocytes and bone-marrow stromal cells were obtained from Clonetics (San Diego, CA). Human primary lymph-node stromal cells were cultured from normal human lymph-node tissue as described³³.

Real-time quantitative PCR

Total RNA from cells or homogenized tissue samples was extracted and reverse transcribed as described⁸. Complementary DNA was quantitatively analysed for the expression of human chemokines and chemokine receptors by the fluorogenic 5'-nuclease PCR assay³⁶ as reported⁸. Specific primers and probes were obtained from Applied Biosystems. Gene-specific PCR products were continuously measured by means of an ABI PRISM 7700 Sequence Detection System (Applied Biosystems) during 40 cycles. For metastasis quantification, specific primers for human *HPRT* which do not cross-react with its mouse counterpart were designed (forward, 5'-TTCCCTGGTCAGGCAGTATAATCC-3'; reverse, 5'-AGTCTGGCTTATATCCAACTTCG-3'). 18S ribosomal RNA was used for normalization.

Tissue samples and immunohistochemistry

Tissue samples of primary tumours from untreated patients ($n = 12$) with invasive ductal

($n = 9$) or lobular ($n = 3$) breast carcinoma were taken, with informed consent, from either diagnostic biopsies or after lumpectomy/mastectomy. Histopathological diagnosis was confirmed for each specimen. Tissue samples of normal human organs were obtained from the National Disease Research Interchange. RNA of various normal human organs was obtained from Clontech. The study was approved by the local ethics committee.

Primary tumours and metastases were routinely fixed with formalin and embedded in paraffin. Antigen retrieval was performed and sections were stained with either anti-human CXCR4 monoclonal antibody (12G5, IgG_{2a}) or IgG_{2a} isotype (R&D Systems) using a standard indirect avidin-biotin horseradish peroxidase method. Colour was developed with diaminobenzidine (DAB) and sections were counterstained with haematoxylin. For staining human CXCR4 in paraffin-embedded or cryo-preserved tissue samples of tumour-bearing SCID mice, 3-amino-9-ethylcarbazole (AEC) was used for colour development.

Actin polymerization assay and microscopy

Actin polymerization was tested as described^{17,37}. Breast cancer cells were incubated either with CXCL12/SDF-1 α or CCL21/6CKine (R&D Systems) at concentrations shown to be optimal in chemotaxis experiments, or with CX₃CL1/fractalkine as negative control. At the indicated time points, cells were fixed, permeabilized and stained in a solution containing paraformaldehyde, lysophosphatidylcholine and fluorescein isothiocyanate (FITC)-labelled phalloidin. Fixed cells were subjected to flow cytometry or analysed by confocal microscopy (Leica DMR). For experiments with adherent cells, breast cancer cells were pre-seeded and incubated with CXCL12, CCL21 or assay buffer (DMEM/0.1% bovine serum albumin (BSA)/12 mM HEPES) for 20 min. Cells were stained with phalloidin and 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) (Molecular Probes).

Chemotaxis and chemoinvasion assays

Migration and invasion was assayed in 24-well cell-culture chambers using inserts with 8- μ m pore membranes as described³⁸. Membranes were pre-coated with fibronectin (2.5–7.5 μ g ml⁻¹) for chemotaxis or Matrigel (28 μ g per insert) and fibronectin for invasion studies. Breast cancer cells were resuspended in chemotaxis buffer (DMEM/0.1% BSA/12 mM HEPES) at 2 or 4 $\times 10^4$ cells per ml. After incubation for 6 or 24 h for chemotaxis or chemoinvasion assays, respectively, cells on the lower surface of the membrane were stained and counted under a light microscope in at least five different fields (original magnification, $\times 200$). Assays were performed in triplicates. Chemokinesis was tested in checkerboard assays and was uniformly negative for both CXCL12 and CCL21.

Proteins from homogenized normal human lung, liver, skin and muscle were extracted in Tris-HCl and protease inhibitor as described³⁹. Conditioned media from human primary bone-marrow or lymph-node stromal cells were generated as described^{16,17,33}.

For neutralization studies, cells were pre-incubated with various concentrations of either anti-human CXCR4 monoclonal antibody (44717.111, IgG_{2b}), anti-human CXCL12 polyclonal antibody (goat IgG) or anti-human CCL21 polyclonal antibody (goat IgG) (all R&D Systems).

In vivo metastasis studies

CB-17 SCID mice (Taconic Farms, Germantown, NY) were injected with MDA-MB-231 breast carcinoma cells either i.v. (10^6 cells) into the tail vein, orthotopically into the inguinal mammary fat pad (10^7 cells). Mice were treated with intraperitoneal injections of a neutralizing anti-human CXCR4 antibody (44717.111, IgG_{2b}) or control IgG_{2b} twice weekly (1 mg per injection). No cytotoxicity for either the anti-human CXCR4 antibody or control IgG_{2b} could be detected in *in vitro* assays. For studies of experimental metastasis, lungs were collected on day 28 and fixed or snap frozen for immunohistochemistry and RNA extraction.

Micro-metastasis was quantified by counting the total tissue area per lung section (D1) and micro-metastasis present in the same area (D2) using a 21-mm² reference grid. The metastatic index was calculated by the ratio D2/D1. For spontaneous metastasis experiments, primary tumours and organs were collected on day 44 after injection. The number of lung colonies was counted by light microscopy (magnification, $\times 40$). The anti-human CXCR4 monoclonal antibody showed no cross-reactivity with murine CXCR4.

Received 15 May 2000; accepted 17 January 2001.

- Nicolson, G. L. Paracrine and autocrine growth mechanisms in tumor metastasis to specific sites with particular emphasis on brain and lung metastasis. *Cancer Metastasis Rev.* **12**, 325–343 (1993).
- Wang, J. M., Deng, X., Gong, W. & Su, S. Chemokines and their role in tumor growth and metastasis. *J. Immunol. Methods* **220**, 1–17 (1998).
- Yeaman, T. J. & Nicolson, G. L. Molecular basis of tumor progression: mechanisms of organ-specific tumor metastasis. *Semin. Surg. Oncol.* **9**, 256–263 (1993).
- Zlotnik, A. & Yoshie, O. Chemokines: a new classification system and their role in immunity. *Immunology* **12**, 121–127 (2000).
- Campbell, J. J. & Butcher, E. C. Chemokines in tissue-specific and microenvironment-specific lymphocyte homing. *Curr. Opin. Immunol.* **12**, 336–341 (2000).
- Butcher, E. C., Williams, M., Youngman, K., Rott, L. & Briskin, M. Lymphocyte trafficking and regional immunity. *Adv. Immunol.* **72**, 209–253 (1999).
- Morales, J. et al. CTACK, a skin-associated chemokine that preferentially attracts skin-homing memory T cells. *Proc. Natl Acad. Sci. USA* **96**, 14470–14475 (1999).
- Honey, B. et al. Cutting edge: the orphan chemokine receptor G protein-coupled receptor-2 (GPR-2, CCR10) binds the skin-associated chemokine CCL27 (CTACK/ALP/ILC). *J. Immunol.* **164**, 3465–3470 (2000).
- Peled, A. et al. Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on

- CXCR4. *Science* **283**, 845–848 (1999).
- Forster, R. et al. CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23–33 (1999).
- Beahrs, O. H. in *Manual for Staging of Cancer* 4th edn (JB Lippincott, Philadelphia, 1992).
- Tashiro, K. et al. Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins. *Science* **261**, 600–603 (1993).
- Bruce, J., Carter, D. C. & Fraser, J. Patterns of recurrent disease in breast cancer. *Lancet* **1**, 433–435 (1970).
- Gunn, M. D. et al. A chemokine expressed in lymphoid high endothelial venules promotes the adhesion and chemotaxis of naive T lymphocytes. *Proc. Natl Acad. Sci. USA* **95**, 258–263 (1998).
- Gunn, M. D. et al. Mice lacking expression of secondary lymphoid organ chemokine have defects in lymphocyte homing and dendritic cell localization. *J. Exp. Med.* **189**, 451–460 (1999).
- Bluel, C. C., Fuhlbrigge, R. C., Casasnovas, J. M., Aiuti, A. & Springer, T. A. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J. Exp. Med.* **184**, 1101–1109 (1996).
- Burger, J. A., Burger, M. & Kipps, T. J. Chronic lymphocytic leukemia B cells express functional CXCR4 chemokine receptors that mediate spontaneous migration beneath bone marrow stromal cells. *Blood* **94**, 3658–3667 (1999).
- Verschueren, H., Van der Taelen, I., Dewit, J., De Brackeleer, J. & De Baetselier, P. Metastatic competence of BW5147 T-lymphoma cell lines is correlated with in vitro invasiveness, motility and F-actin content. *J. Leukoc. Biol.* **55**, 552–556 (1994).
- Bluel, C. C. et al. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
- Hromas, R. et al. Isolation and characterization of Exodus-2, a novel C-C chemokine with a unique 37-amino acid carboxyl-terminal extension. *J. Immunol.* **159**, 2554–2558 (1997).
- Hedrick, J. A. & Zlotnik, A. Identification and characterization of a novel beta chemokine containing six conserved cysteines. *J. Immunol.* **159**, 1589–1593 (1997).
- Campbell, J. J. et al. Chemokines and the arrest of lymphocytes rolling under flow conditions. *Science* **279**, 381–384 (1998).
- Luboshits, G. et al. Elevated expression of the CC chemokine regulated on activation, normal T cell expressed and secreted (RANTES) in advanced breast carcinoma. *Cancer Res.* **59**, 4681–4687 (1999).
- Belperio, J. A. et al. CXC chemokines in angiogenesis. *J. Leukoc. Biol.* **68**, 1–8 (2000).
- Vicari, A. P. et al. Antitumor effects of the mouse chemokine 6CKine/SLC through angiostatic and immunological mechanisms. *J. Immunol.* **165**, 1992–2000 (2000).
- Ma, Q. et al. Impaired B-lymphopoiesis, myelopoiesis, and deranged cerebellar neuron migration in CXCR4- and SDF-1-deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 9448–9453 (1998).
- Tachibana, K. et al. The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* **393**, 591–594 (1998).
- Nagasawa, T. et al. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
- Zou, Y. R., Kottmann, A. H., Kuroda, M., Taniuchi, I. & Littman, D. R. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature* **393**, 595–599 (1998).
- Vassileva, G. et al. The reduced expression of 6CKine in the plt mouse results from the deletion of one of two 6CKine genes. *J. Exp. Med.* **190**, 1183–1188 (1999).
- Mohle, R. et al. Functional response of leukaemic blasts to stromal cell-derived factor-1 correlates with preferential expression of the chemokine receptor CXCR4 in acute myelomonocytic and lymphoblastic leukaemia. *Br. J. Haematol.* **110**, 563–572 (2000).
- Mohle, R., Failenschmid, C., Bautz, F. & Kanz, L. Overexpression of the chemokine receptor CXCR4 in B cell chronic lymphocytic leukemia is associated with increased functional response to stromal cell-derived factor-1 (SDF-1). *Leukemia* **13**, 1954–1959 (1999).
- Arai, J., Yasukawa, M., Yakushiji, Y., Miyazaki, T. & Fujita, S. Stromal cells in lymph nodes attract B-lymphoma cells via production of stromal cell-derived factor-1. *Eur. J. Haematol.* **64**, 323–332 (2000).
- Koshiba, T. et al. Expression of stromal cell-derived factor 1 and CXCR4 ligand receptor system in pancreatic cancer: a possible role for tumor progression. *Clin. Cancer Res.* **6**, 3530–3535 (2000).
- Hendrix, C. W. et al. Pharmacokinetics and safety of AMD-3100, a novel antagonist of the CXCR4 chemokine receptor, in human volunteers. *Antimicrob. Agents Chemother.* **44**, 1667–1673 (2000).
- Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 7276–7280 (1991).
- Howard, T. H. & Meyer, W. H. Chemotactic peptide modulation of actin assembly and locomotion in neutrophils. *J. Cell Biol.* **98**, 1265–1271 (1984).
- Saiki, I., Murata, J., Nakajima, M., Tokura, S. & Azuma, I. Inhibition by sulfated chitin derivatives of invasion through extracellular matrix and enzymatic degradation by metastatic melanoma cells. *Cancer Res.* **50**, 3631–3637 (1990).
- Hujanen, E. S. & Terranova, V. P. Migration of tumor cells to organ-derived chemoattractants. *Cancer Res.* **45**, 3517–3521 (1985).

Supplementary information is available on Nature's World-Wide web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. M. Barcenas for technical assistance; S. Ulloa and K. Smith for help procuring tissue samples; and D. J. Ruiter, I. J. Fidler and D. Schadendorf for the melanoma cell lines. We are grateful to R. de Waal Malefyt for establishing the quantitative RT-PCR at the DNAX Research Institute, and H. Kanzler, T. Hauser and L. Bakker for critical comments on the manuscript. B.H. was supported by a grant from the Deutsche Forschungsgemeinschaft. DNAX Research Institute is supported by the Schering-Plough Corporation.

Correspondence and requests for materials should be addressed to A.Z. (e-mail: azlotnik@eosbiotech.com).

Flooding of Ganymede's bright terrains by low-viscosity water-ice lavas

Paul M. Schenk^{*}, William B. McKinnon[†], David Gwynn[‡] & Jeffrey M. Moore[§]

^{*} Lunar and Planetary Institute, Houston, Texas 77058, USA

[†] Department of Earth and Planetary Sciences, and McDonnell Center for the Space Sciences, Washington University, Saint Louis, Missouri 63130, USA

[‡] Department of Geography, University of California, Los Angeles, California 90095, USA

[§] NASA Ames Research Center, Moffett Field, California 94035, USA

Large regions of the jovian moon Ganymede have been resurfaced, but the means has been unclear^{1,2}. Suggestions have ranged from volcanic eruptions of liquid water^{3–5} or solid ice⁶ to tectonic deformation^{7–9}, but definitive high-resolution morphological evidence has been lacking. Here we report digital elevation models of parts of the surface of Ganymede, derived from stereo pairs combining data from the Voyager and Galileo spacecraft, which reveal bright, smooth terrains that lie at roughly constant elevations 100 to 1,000 metres below the surrounding rougher terrains. These topographic data, together with new images that show fine-scale embayment and burial of older features¹⁰, indicate that the smooth terrains were formed by flooding of shallow

structural troughs by low-viscosity water-ice lavas. The oldest and most deformed areas (the 'reticulate' terrains) in general have the highest relative elevations, whereas units of the most common resurfaced type—the grooved terrain—lie at elevations between those of the smooth and reticulate terrains. Bright terrain, which accounts for some two-thirds of the surface, probably results from a continuum of processes, including crustal rifting, shallow flooding and groove formation. Volcanism plays an integral role in these processes, and is consistent with partial melting of Ganymede's interior^{11,12}.

Ganymede's surface can be divided into older, heavily cratered, dark terrain (~34% by area, based on a global 2-km-resolution mosaic produced by P.M.S.) and younger, resurfaced, bright terrain (~66%). The ages¹³ and nature of the resurfacing, if determined, would place important constraints on Ganymede's evolution (for example, on the role of tidal heating, which has been proposed as a trigger for internal melting and resurfacing on Ganymede^{11,12}). Voyager-based analyses concluded that bright terrain, including both smooth and grooved terrain units, probably formed by resurfacing of shallow structural graben, troughs or rifts in Ganymede's dark, ancient crust^{1,2,4}. Although volcanism was inferred, no direct evidence was observed to indicate how the resurfacing occurred. At least 18 caldera-like features were identified globally^{7,14}, but their role in resurfacing was unclear.

Except for a high-standing ridged unit in one of the calderas^{10,15}, sought-after evidence for individual lava flows or eruption centres has not been resolved in high-resolution (<300 m per pixel) Galileo images to date. This suggests that flow fronts have been destroyed by fracturing, impact erosion or mass wasting¹⁶, or are too low in relief

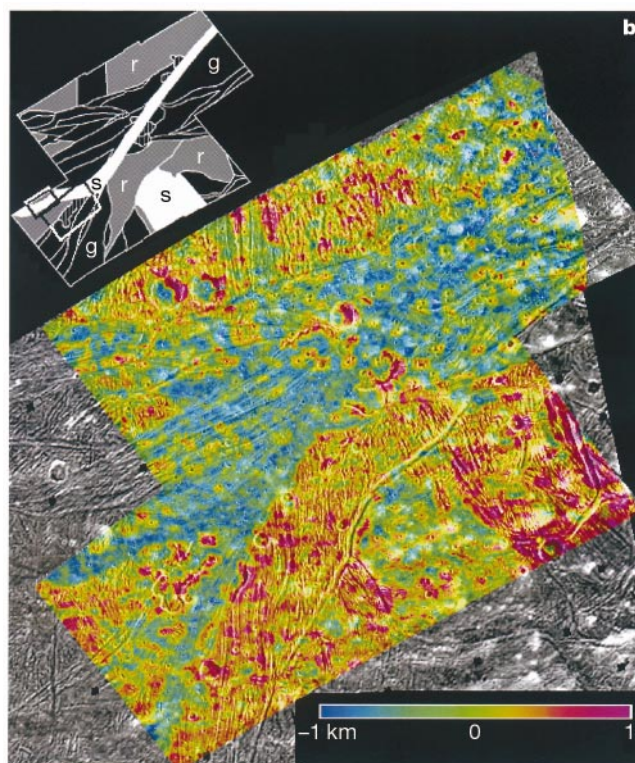
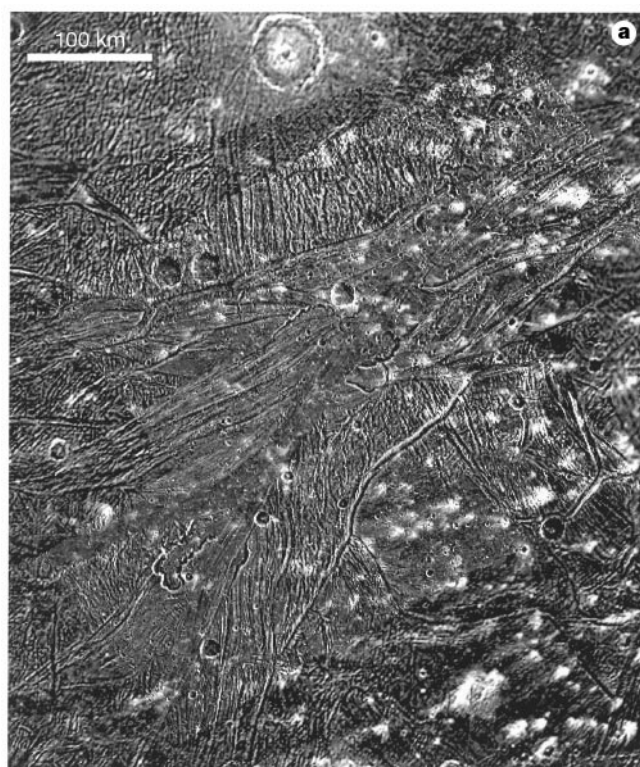


Figure 1 Bright terrain units and topography within Sippar Sulcus, Ganymede. The three dominant structural styles—grooved terrain, smooth terrain and reticulate terrain—are represented. **a**, Galileo image mosaic obtained with a resolution of 180 m per pixel, superposed on regional Voyager images of resolution 1 km per pixel. The swath of smooth terrain crosses the scene diagonally from upper right to centre left. Irregularly shaped enclosures are interpreted as calderas^{14,15}. The numerous bright patches are due to secondary impacts from the large crater Osiris, ~600 km to the ESE of the centre of the

mosaic (30° S, 184° W). Images used are FDS 20637.41, 20637.53 and 20637.56 (Voyager) and 394532413–39453278 (Galileo). **b**, DEM of topography of the same scene. Relative elevation values have been colour-coded and merged with the Galileo mosaic. The DEM has a horizontal resolution and sample spacing of 7 km and 1 km, respectively; vertical resolution is 0.2 km. The inset shows a geologic map highlighting areas of grooved terrain (g, black), reticulate terrain (r, grey), smooth terrain (s, white), calderas (hatched), and locations for Fig. 3a and b (upper and lower box, respectively).

to be resolved by Galileo, or that bright terrain simply did not form by extrusion of lavas. The lack of clear morphological evidence for volcanism, together with the discovery of pervasive faulting in grooved terrains in Uruk Sulcus¹⁷, has led to (or revived⁷) an alternative model: bright terrain might be mostly formed by pervasive tectonic disruption and brightening of older terrains, a process referred to as "tectonic resurfacing"^{8,9,18}.

Regional differences in elevation between resurfaced terrains of different ages can help to distinguish emplacement mechanisms. Although crosscutting relationships have long suggested that bright terrain may be lower than dark terrain^{1–5}—in that bright terrain is nearly always tectonically confined—offsets were never quantitatively determined. As of this writing, targeted Galileo stereo observations and derived topographic maps (digital elevation models, or DEMs) of bright terrain show only local-scale features, and do not include terrain boundaries¹⁹. To map the relative topography of resurfaced units, we have combined Voyager and Galileo images to produce eight regional-scale stereo pairs. DEMs were generated using stereogrammetric autocorrelation software²⁰. The areal coverage and associated high-resolution CCD (charge-coupled device) images in these stereo mosaics offer a notable improvement over low-resolution Voyager-only stereo pairs²¹.

We have identified two sites that are best suited for investigation of the roles of volcanism and tectonism. The Sippar Sulcus site (Fig. 1) features numerous crosscutting bright terrain units and calderas. Geologic mapping by us and Wilhelms²² indicates that rugged, reticulate terrains (which are intensely deformed by multiple groove sets) are the oldest units in this region. Crosscutting these are numerous younger (less cratered) grooved terrain units, characterized by varying degrees of relief interpreted as variations in intensity of tectonic deformation. Although relative ages of grooved units are difficult to correlate regionally, the oldest in Sippar Sulcus tend to be the most intensely tectonically deformed (that is, they have the greatest areal density of—or highest degree of relief across—grooves). The last non-impact geologic events in this region were the formation of a number of cusped, open-ended calderas^{7,14} (seven can be seen in Fig. 1a) and a distinctive 15–25-km wide swath of smooth terrain; all crosscut grooved terrain units.

Our DEM of Sippar Sulcus reveals that reticulate, grooved and smooth terrains lie at characteristically different relative elevations (Fig. 1b). The oldest and most strongly deformed units (reticulate terrain) are the highest standing. The youngest and least deformed

unit (the smooth terrain swath) is the lowest, lying ~100–500 m below older, crosscut lanes of grooved terrain, and ~800–1,100 m below older, reticulate terrains (Fig. 1b). The elevations of grooved terrain units within Sippar Sulcus are variable, ranging between those of reticulate terrain and the youngest smooth terrain. The trapezoidal patch of smooth (or very lightly grooved) terrain at the southern margin of the DEM lies at an intermediate elevation. Although elevations of adjacent units are different, the elevation along the swath of smooth terrain is, to the limit of DEM vertical resolution, constant (or at most, gently sloping, with no undulations or breaks in slope) across our map area, a distance of over 500 km.

These relationships are further illuminated at the intersection of Erech Sulcus and northern Sippar Sulcus (Fig. 2a). The intensely grooved terrain within Erech Sulcus mostly lies at elevations comparable to adjoining dark terrain. These units are truncated by the northern margin of Sippar Sulcus, which is defined by a lane of smooth terrain ~8 km wide and lying ~750 m below adjacent dark terrain (Fig. 2b). Interior portions of Sippar Sulcus here consist of truncated packets of grooved material crosscut by topographically lower units of younger smooth material. Elevations of these grooved terrain units are intermediate between those of smooth terrain and the nearby dark terrains. Thus we observe the same general correlation between elevation and degree of tectonic deformation (that is, groove relief and density) at two distinct sites.

Uniformly low-standing topography of smooth terrains, especially in comparison to the variable topography of adjacent older grooved and reticulate terrain units, strongly suggests down-dropping and flooding to an equipotential surface. In addition, the presence of only occasional bright lineations within the swath of smooth terrain suggests that tectonism (or at least deformation capable of destroying ridges hundreds of metres high) has not been an important process within this terrain. The embayment of individual older ridges and grooves along the margin of this swath²², confirmed in Galileo images¹⁰ (Fig. 3a), is also characteristic of flooding of low-lying topography, and is inconsistent with emplacement by tectonic disruption. Similar embayment of ridges is observed at several sites along a swath of relatively young smooth terrain near the south pole^{5,14} (see, for example, figure 1 in ref. 5) and elsewhere²², indicating that smooth terrains globally may have similar origins.

Tectonic resurfacing^{8,9,18}, in contrast, should lead to intense

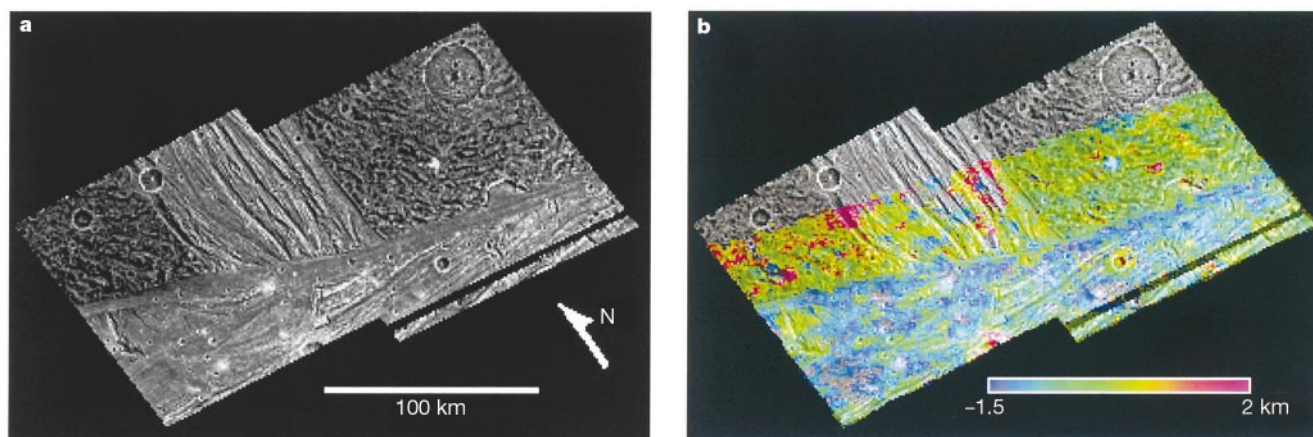


Figure 2 Terrain units and topography at the intersection of Erech Sulcus and northern Sippar Sulcus, Ganymede. **a**, Galileo image mosaic obtained with a resolution of 140 m per pixel. Erech Sulcus is the 75-km-wide band of grooved terrain extending between the two blocks of older, dark terrain. Portions of Sippar Sulcus lie to the south, trending at high angles to Erech Sulcus. Note the caldera at right, truncated by the narrow lane of smooth

terrain. **b**, DEM of topography of the same scene. Relative elevation values have been colour-coded and merged with the mosaic. Horizontal resolution and sample spacing are 7 and 1 km, respectively; vertical resolution is 0.24 km. Scene is centred at 16° S, 177° W, roughly 500 km north of the centre of Fig. 1. Images used are FDS 20637.41 (Voyager), and 394533065 and 394533078 (Galileo).

groove formation within 'new' terrains and some preservation of broad-scale relative elevation differences of older crosscut and faulted terrains. Topographic downdropping is logically associated with tectonic resurfacing^{9,18}, but the lowest units here are the least deformed and do not preserve vestiges of pre-existing relative elevation differences. We also find no tectonic evidence for shear (previously hypothesized⁷) within or adjacent to the smooth swath (Figs 1a and 3); nor can shear obviously explain the topography or topology of a depressed, dog-legged, smooth terrain swath. We also find no evidence that the two sides of the smooth terrain swath were originally together, as might be expected if the terrain formed by crustal spreading. We therefore conclude that volcanic resurfacing of shallow structural troughs (see, for example, ref. 4) is indeed primarily responsible for the formation of smooth terrain (at least in Sippar Sulcus and near the south pole¹⁴).

Tectonism remains crucial to Ganymede's geology, however. Extensional tectonism is responsible for downdropping blocks of dark terrain in the first place, and for extending both dark terrain and bright, resurfaced terrain to create lanes of grooved and reticulate terrain^{3,4}. Nor do our observations contradict those at Uruk Sulcus, where grooved terrain formed entirely by tectonic extension (and some shear)⁹ from what Voyager identified as smooth terrain. This 'smooth' terrain, when seen at high resolution (75 m per pixel), is bright and flat-topped, with a structural imprint of subparallel fine crevasses and simple graben (see, for example, figure 5 in ref. 9). The important point is that this precursor bright terrain is flat overall, and only modestly tectonized; logically, it was originally emplaced as a volcanic unit. Likewise, the smooth terrains in Figs 1–3 are at most lightly tectonized (grooved).

The relative continuity of elevation and the lack of prominent volcanic textures or relief within smooth terrains studied here imply that volcanic resurfacing materials in smooth terrains have a relatively low viscosity. Based on terrestrial volcanic experience, resurfacing by solid-state flow of ice would be expected to produce flows of limited extent and substantial relief¹⁴. Such relief might go undetected in our DEMs, but high-resolution images (for example, Fig. 3) suggest that the relief of flow units within smooth terrain is very subtle and may be tens of metres or less. In contrast, the high-standing ridged deposit within the largest Sippar Sulcus caldera (Fig. 3b) has been inferred to be a relatively viscous flow^{10,15}.

Similarly high-standing (and sometimes ridged) materials exist within the other calderas^{10,14,15}, but no similar morphological evidence for viscous flow outside of caldera margins has been identified to date. Our DEMs (and photoclinometric profiles¹⁴) indicate that caldera floor elevations are also variable, ranging between those of surrounding grooved terrains and adjacent smooth terrain (Figs 1b, 2b and 3b).

Resurfacing of smooth terrain on Ganymede may be in some respects analogous to that of the lunar mare, which are characterized by shallow flooding of topographic lows with low-viscosity lavas and a general lack of volcanic edifices. The lunar volcanic style is plausibly related to, among other things, the negative buoyancy of basaltic magma with respect to anorthosite-rich crust, and infrequent, high-volume eruptions²³. On Ganymede, our data confirm that smooth terrain deposits are similarly confined to low-lying areas, indicating that smooth terrain magmas are negatively buoyant with respect to Ganymede crust as well (as first argued by Shoemaker *et al.*³). The long-standing problem of how a relatively dense, aqueous magma evolves out of Ganymede's icy mantle^{1,2,5,6} remains, but it is possible that tidal heating within the convecting portion of the upper mantle led to ascent of buoyant masses of partially molten ice, providing both the extensional stress and liquid for resurfacing¹².

Our observations suggest a relatively simple model, in which each successive unit of bright terrain forms initially as a structural trough and is subsequently flooded by low-viscosity aqueous lava, which then freezes. Caldera formation may precede this structural downdropping, and certainly precedes the flooding that forms smooth terrain (ridged caldera floor material represents a higher-viscosity volcanic stage). Extension and groove formation can then follow smooth terrain flooding. Extension may further lower topographic levels⁹, but may be counteracted by isostatic adjustments owing to the high heat flows associated with groove formation²⁴. The general correlation of elevation and deformation is consistent with decreasing intensity of groove formation over time. As most bright terrain on Ganymede is grooved, only the most recently formed smooth units have survived (relatively) undisrupted.

We find volcanic resurfacing to be an integral part of the process of bright terrain formation on Ganymede in at least two geographically distinct regions. High-resolution stereo observations

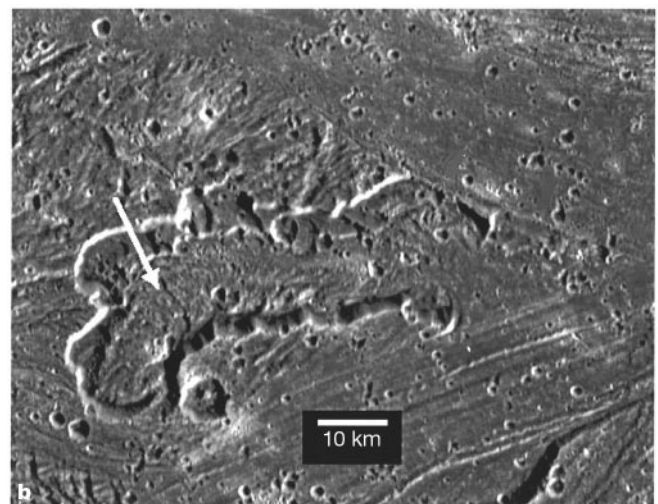
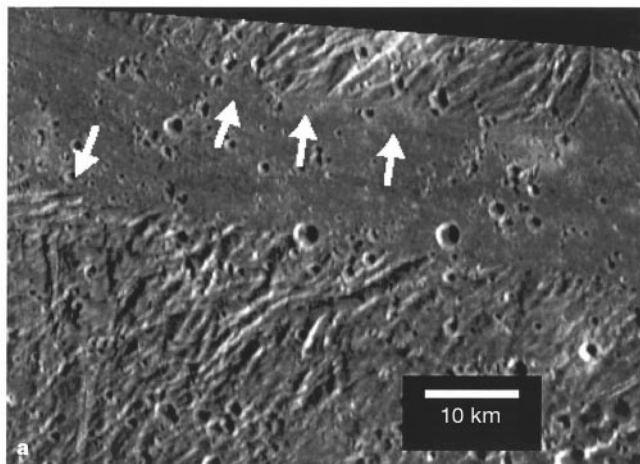


Figure 3 High-resolution views of Sippar Sulcus. **a**, Embayment of ridges and troughs (arrows) in older reticulate terrain^{10,22} by smooth terrain, indicative of flooding by low-viscosity materials. **b**, Irregularly shaped caldera and high-standing interior; the latter is interpreted to be a viscous flow^{10,15}. Our DEM (Fig. 1b) shows that the elevation of the westernmost caldera floor material (arrow) is comparable to adjacent grooved unit but

decreases towards the east (right), where it is similar to nearby, lower-lying smooth terrains. Note the general lack of lineations within smooth terrain, which extends across the upper half of the image and crosscuts a similar but grooved unit at lower right. Both views are portions of Galileo image 394532478; resolution is 180 m per pixel. North is up. See Fig. 1b for locations.

of terrain boundaries taken during orbit G28 in May 2000, and transmitted to Earth over the following months, should indicate whether the topographic relationships we have described characterize bright terrains generally. It remains for future studies to unravel the global distribution, timing and relative importance of volcanism relative to tectonism on Ganymede. □

Methods

To measure parallax, the stereogrammetry technique^{30,21} relies on matching of albedo patterns within a sample area centred on each pixel (this sample area is equivalent to the 'footprint' of the elevation measurement, 3–13 times the image resolution in these maps, depending on scene content). Some features (for example, bright crater ejecta) appear different in the stereo images due to photometric effects, and can contribute noise to the final DEM. Because no global topographic datum exists for Ganymede, our DEMs measure local relief only; they are not controlled with respect to the centre of Ganymede. Limb images indicate that there is little pronounced regional-scale topographic relief, consistent with low relief generally³. Assuming that the mean slope across each site is negligible, our stereo pairs were semi-controlled by selecting registration (or tie) points within geologic units assumed to lie at similar elevations. For the Sippar Sulcus site, tie points were located within older reticulate terrain. In the Erech Sulcus site, tie points were selected in older dark terrain.

Received 7 August; accepted 14 December 2000.

- McKinnon, W. B. & Parmentier, E. M. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 718–763 (Univ. Arizona Press, Tucson, 1986).
- Squires, S. W. & Croft, S. K. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 293–341 (Univ. Arizona Press, Tucson, 1986).
- Shoemaker, E. M., Lucchitta, B. K., Plescia, J. B., Squires, S. W. & Wilhelms, D. E. in *Satellites of Jupiter* (ed. Morrison, D.) 435–520 (Univ. Arizona Press, Tucson, 1982).
- Parmentier, E. M., Squires, S. W., Head, J. W. & Allison, M. L. The tectonics of Ganymede. *Nature* **295**, 290–293 (1982).
- Allison, M. L. & Clifford, S. M. Ice-covered water volcanism on Ganymede. *J. Geophys. Res.* **92**, 7865–7876 (1987).
- Kirk, R. L. & Stevenson, D. J. Thermal evolution of a differentiated Ganymede and implications for surface features. *Icarus* **69**, 91–134 (1987).
- Lucchitta, B. K. Grooved terrain on Ganymede. *Icarus* **44**, 481–501 (1980).
- Head, J. W., Pappalardo, R. T., Collins, G. & Greeley, R. Tectonic resurfacing on Ganymede and its role in the formation of grooved terrain. *Lunar Planet. Sci. Conf. [CD-ROM] XXIX*, 535 (1997).
- Pappalardo, R. T. *et al.* Grooved terrain on Ganymede: First results from Galileo high-resolution imaging. *Icarus* **135**, 276–302 (1998).
- Kay, J. E. & Head, J. W. Geologic mapping of the Ganymede G8 Calderas region: Evidence for cryovolcanism. *Lunar Planet. Sci. Conf. [CD-ROM] XXX*, 1103 (1999).
- Showman, A. P. & Malhotra, R. Tidal evolution into the Laplace resonance and the resurfacing of Ganymede. *Icarus* **127**, 93–111 (1997).
- Showman, A. P., Stevenson, D. J. & Malhotra, R. Coupled orbital and thermal evolution of Ganymede. *Icarus* **129**, 367–383 (1997).
- Zahnle, K., Dones, L. & Levison, H. F. Cratering rates on the Galilean satellites. *Icarus* **136**, 202–222 (1998).
- Schenk, P. M. & Moore, J. M. Volcanic constructs on Ganymede and Enceladus: Topographic evidence from stereo images and photogrammetry. *J. Geophys. Res.* **100**, 19009–19022 (1995).
- Head, J. W. *et al.* Cryovolcanism on Ganymede: Evidence in bright terrain from Galileo solid state imaging data. *Lunar Planet. Sci. Conf. [CD-ROM] XXIX*, 1666 (1998).
- Moore, J. M. *et al.* Mass movement and landform degradation on the icy Galilean satellites: Results of the Galileo nominal mission. *Icarus* **140**, 294–312 (1999).
- Belton, M. J. S. *et al.* Galileo's first images of Jupiter and the Galilean satellites. *Science* **274**, 377–385 (1996).
- Collins, G. C., Head, J. W. & Pappalardo, R. T. Formation of Ganymede grooved terrain by sequential extensional episodes: Implications of Galileo observations for regional stratigraphy. *Icarus* **135**, 345–359 (1998).
- Giese, B. *et al.* The local topography of Uruk Sulcus and Galileo Regio obtained from stereo images. *Icarus* **135**, 303–316 (1998).
- Schenk, P. M. & Moore, J. M. Stereo topography of the south polar region of Mars: Volatile inventory and Mars Polar Lander landing site. *J. Geophys. Res.* **105**, 24529–24546 (2000).
- Schenk, P. M., Jones, K. & Fessler, B. The 400 meters: Stereo topographic mapping of the bright terrain-dark terrain boundary on Ganymede. *Bull. Am. Astron. Soc.* **28**, 1074 (1996).
- Wilhelms, D. E. Geologic map of the Osiris (Jg-12) and Apsu Sulci (Jg-13) Quadrangles of Ganymede. *US Geol. Surv. Misc. Invest. Ser. Map I-2442* (1997).
- Head, J. W. & Wilson, L. Lunar mare volcanism—Stratigraphy, eruption conditions, and the evolution of secondary crusts. *Geochim. Cosmochim. Acta* **56**, 2155–2175 (1992).
- Dombard, A. J. & McKinnon, W. B. Formation of grooved terrain on Ganymede: Extensional instability mediated by cold, superplastic creep. *Icarus* (submitted).

Acknowledgements

We thank K. Jones for image processing during the early phases of this project, L. Prockter and especially R. Kirk for comments on the manuscript, and NASA's Jovian Systems Data Analysis Program and the Planetary Geology & Geophysics Program for support.

Correspondence and requests for materials should be addressed to P.M.S. (e-mail: schenk@pi.usra.edu).

Coherent manipulation of semiconductor quantum bits with terahertz radiation

B. E. Cole[†], J. B. Williams^{*}, B. T. King^{*}, M. S. Sherwin^{*} & C. R. Stanley[‡]

^{*} Institute for Quantum Engineering, Science and Technology (iQUEST) and Department of Physics, University of California, Santa Barbara, California 93106, USA

[‡] Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, G12 8QQ, UK

Quantum bits^{1–5} (qubits) are the fundamental building blocks of quantum information processors, such as quantum computers⁶. A qubit comprises a pair of well characterized quantum states that can in principle be manipulated quickly compared to the time it takes them to decohere by coupling to their environment⁷. Much remains to be understood about the manipulation and decoherence of semiconductor qubits. Here we show that hydrogen-atom-like motional states of electrons bound to donor impurities in currently available semiconductors can serve as model qubits. We use intense pulses⁸ of terahertz radiation to induce coherent, damped Rabi oscillations^{9,10} in the population of two low-lying states of donor impurities in GaAs^{11–13}. Our observations demonstrate that a quantum-confined extrinsic electron in a semiconductor can be coherently manipulated like an atomic electron, even while sharing space with $\sim 10^5$ atoms in its semiconductor host. We anticipate that this model system will be useful for measuring intrinsic decoherence processes, and for testing both simple and complex manipulations of semiconductor qubits.

An electron bound to a 'shallow donor' (for example, a silicon or a sulphur atom) in GaAs has a hydrogenic character with a large Bohr radius of ~ 10 nm (ref. 13). Photoconductivity has been extensively used for continuous-wave (CW) spectroscopy of the energy level structure of donor systems^{13,14}. The hydrogenic $1s-2p$ ($m = +1$) (hereafter, $1s-2p^+$) transition lies in the terahertz frequency range and may be tuned from 1 THz to over 5 THz by the application of a magnetic field of just a few tesla. Here m is the quantum number associated with the projection of the orbital angular momentum onto the magnetic field axis. At magnetic fields relevant to this work, the $2p^+$ state is resonant with ionized states. The three processes characterizing the photoconductive response of the $1s-2p^+$ system are illustrated in Fig. 1a. Most electrons promoted to the $2p^+$ state do not return to the ground state directly, but are first ionized to the conduction band, where they contribute to the macroscopic conductivity of the sample.

Previous studies¹⁵ have determined the lifetime of the $2p^+$ level to be not more than 300 ps under weak terahertz driving conditions. Transition linewidths indicate a dephasing time shorter than this, due to inhomogeneous broadening. Larsen has shown¹⁶ that inhomogeneous broadening arises from random electric fields associated with residual compensating impurities. Planken *et al.* observed¹⁷ the free-induction decay (FID) of the $1s-2p^+$ n-GaAs transition by probing with weak terahertz pulses¹⁷. If the $1s$ and $2p^+$ states of donors are mapped onto up and down states of fictitious spins in a magnetic field, the weak pulses in FID experiments rotate the fictitious spins through an angle $\theta \ll 1$. In the present work, intense terahertz pulses are used to coherently rotate the fictitious spins through angles $\theta \gg 2\pi$. Such arbitrary 1-bit rotations are a requirement for quantum information processing⁷.

[†] Present address: Toshiba Research Europe Limited, Cambridge Research Laboratory, 260 Cambridge Science Park, Milton Road, Cambridge, CB4 0WE, UK.

Previous experiments have reported excitonic Rabi flopping in GaAs quantum wells^{18,19} driven by short pulses of intense interband radiation. The intrinsic crystal excitations that were pumped by these experiments are delocalized modes of the solid, which are not easily mapped onto localized two-state systems suitable for use as qubits.

The design of the sample that we investigated is illustrated in Fig. 1b. The sample was driven by intense tunable terahertz radiation provided by the University of California, Santa Barbara (UCSB) free-electron lasers²⁰. A fixed frequency of 2.52 THz was chosen for this experiment, which coincides with a pass band in the atmospheric transmission. The $1s-2p^+$ transition was tuned to the

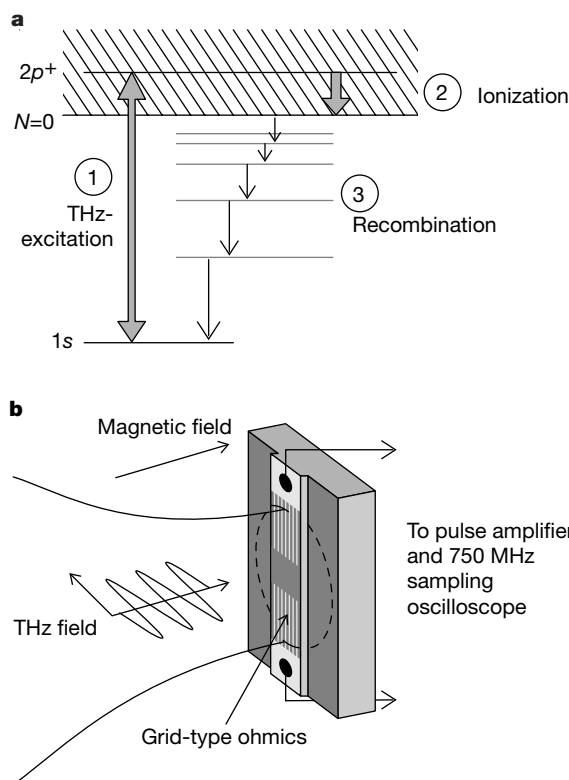


Figure 1 Schematic drawing of terahertz (THz) photoconducting processes and sample. **a**, Processes. (1) After the 'switch-on' of the terahertz field E_{THz} , the bound electron population oscillates between the ground ($1s$) and excited ($2p^+$) states. (2) Electrons left in the $2p^+$ state at the end of the terahertz pulse are ionized into the conduction band continuum of conduction band states associated with the lowest ($N = 0$) Landau level. Thus, the transient increase in conductivity is proportional to the number of electrons excited out of the $1s$ state by the terahertz pulse. (3) The free conduction band electrons recombine with donor ions. **b**, Sample geometry. The sample was fabricated from a 15- μm -thick GaAs layer grown by molecular beam epitaxy on an semi-insulating GaAs substrate²⁵. The electron density was $2.8 \times 10^{14} \text{ cm}^{-3}$ with mobility $140,000 \text{ V}^{-1} \text{ s}^{-1} \text{ cm}^2$ at 77 K (ref. 26). Quasi-CW photoconductivity spectra indicate that a single donor species (sulphur) dominates. The sample had a 20-nm Si-doped ($\sim 10^{18} \text{ cm}^{-3}$) cap layer to facilitate the formation of ohmic contacts (AuNiGe), which was etched off the remaining sample area after alloying the contacts. The active region of the sample ($100 \times 100 \mu\text{m}$) was defined by wet etching and the contacts. This active region was much smaller than the size of the terahertz focus ($500 \mu\text{m}$ full-width at half-maximum, FWHM), ensuring uniform illumination. To minimize distortions of the terahertz fields by the ohmic contacts, these were made of ten equally spaced 10- μm AuNiGe strips perpendicular to E_{THz} . The back of the sample had an impedance-matched anti-reflection coating to eliminate internal reflections²⁷. The sample was placed in the bore of a superconducting solenoid, and immersed in liquid helium at $\sim 2 \text{ K}$. The sample was connected directly to a 40 Ω coaxial cable which fed through a 'bias-tee' into a broadband pulse amplifier and a 750 MHz digital oscilloscope. The sample was biased with an electric field of 1 kV m^{-1} , below the threshold for impact ionization.

laser frequency with a magnetic field of 3.5 T. In order to generate the short terahertz pulses required for this work, light-activated semiconductor switches were used to 'slice' short segments from the 2–6- μs -long pulses from the free-electron laser^{8–21}. Figure 2a shows the energy in a sliced pulse as a function of its duration. From the linear relationship between pulse energy and duration, we deduce that the sliced pulses have a 'flat top' after a rise time which is too small ($< 4 \text{ ps}$) to extract reliably from these data.

In Fig. 2b we show the temporal profile of the photocurrent transient. The exponential decay of the photocurrent is determined by the electron recapture rate, a constant sample parameter for the purposes of this experiment. Excited electrons are known to ionize within the first nanosecond after the end of the terahertz pulse¹⁵. The integrated photocurrent is thus proportional to the total fraction of electrons excited out of the $1s$ ground state at the end of the terahertz pulse.

We plot the integrated photocurrent against pulse duration in Fig. 3a for a range of terahertz fields. The magnetic field was set at 3.56 T, corresponding to the on-resonance condition at the highest terahertz fields. One to two cycles of Rabi-oscillation are clearly visible. The oscillation frequency increases with increasing terahertz field, as expected. However, the rate at which the oscillations are damped also increases, resulting in a number of cycles of oscillation that is roughly independent of terahertz field.

The magnetic field for which the photocurrent rise-time is a maximum corresponds to the on-resonance condition. A contour plot of photocurrent as a function of pulse duration and magnetic field is shown in Fig. 3b for a terahertz field strength of $E_{\text{THz}} = 2.1 \times 10^4 \text{ V m}^{-1}$; the resonant magnetic field is indicated by the upper arrow. We note that the resonant magnetic field in the presence of the strong terahertz driving field is higher than

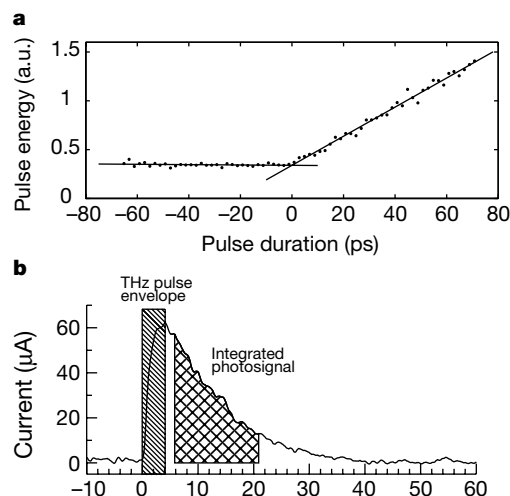


Figure 2 Terahertz pulse shape and photoconductive response. **a**, Pulse energy versus pulse duration. The < 50 -ps pulses used in these experiments were 'sliced' from a longer pulse using semiconductor switches activated by intense ~ 150 -fs near-infrared pulses from a regeneratively amplified Ti:sapphire laser. The terahertz pulses are too short for a real-time measurement of their shape. Each point in this figure is proportional to the pulse energy, as detected by a 4 K InSb hot-electron bolometer, tuned with a fixed magnet, with a time constant of 200 ns. The pulse duration is defined as the time delay between the near-infrared pulses which respectively turn on and off the sliced pulse. The slopes and intercepts of the two lines were deduced by a least-squares fit to the data. The linear increase of pulse energy with pulse duration is consistent with the expected 'flat-topped' pulse shape. **b**, The time dependence of the photocurrent transient due to a weak 4-ns terahertz pulse. The repetition rate of the terahertz pulses was under 10 Hz, ensuring that the system returned to equilibrium after each transient. The prepulse power with switches inactive was 5×10^6 smaller than the power during the sliced pulse, and had no noticeable effect on the sample.

that found from linear photocurrent spectra (that is, in the weak-terahertz limit). The resonant magnetic field increases from 3.51 T at very low E_{THz} (see inset on left side of Fig. 3b), to 3.54 T at $E_{\text{THz}} = 2.1 \times 10^4 \text{ V m}^{-1}$ to 3.56 T at the highest $E_{\text{THz}} = 4.9 \times 10^4 \text{ V m}^{-1}$. This shift is consistent with an a.c. Stark effect¹⁰, wherein coupling to higher states results in level repulsion, lowering

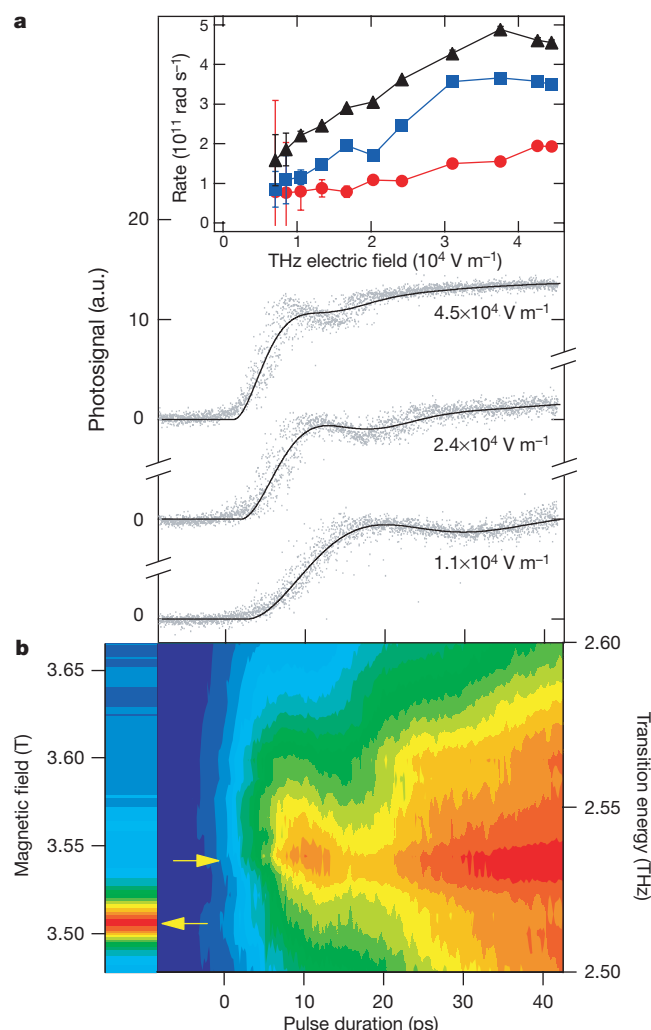


Figure 3 Rabi oscillations in photocurrent. **a**, Photocurrent versus pulse duration for terahertz fields $E_{\text{THz}} = 4.5, 2.4$ and $1.1 \times 10^4 \text{ V m}^{-1}$. Each data point corresponds to a single sliced pulse. The terahertz field inside the sample was estimated from the measured energy of the microsecond-long pulses of the free-electron laser, taking into account losses in the pulse slicer, cryostat windows, and the Fresnel coefficient of the GaAs/air interface. The uncertainty of the absolute scale of E_{THz} is $\pm 50\%$, but uncertainty in relative values of E_{THz} is negligible. The plots are offset for clarity. Black curves are fits to analytic solutions for $1 - \rho_{11}(t)$, solved in the rotating wave approximation assuming resonant excitation. Defining the turn-on time to be $t = 0$, the initial conditions are: $\rho_{11}(0) = 1, \rho_{12}(0) = \rho_{21}(0) = \rho_{22}(0) = 0$. Since the $2p^+ - 1s$ recombination rate γ_1 is expected to be much smaller than γ_2 or γ_3 , we set $\gamma_1 = 0$. At the highest E_{THz} , the measured signal versus pulse duration turns on more slowly than the model predicts, and the depth of the oscillation is smaller in the model than in the data. The discrepancy in rise times at the highest intensities could be explained if the experimental pulse has a rise time of a few picoseconds, rather than the instantaneous turn-on assumed in the theoretical model. Inset, fit parameters versus E_{THz} : Rabi frequency Ω_R (triangles), dephasing rate γ_2 (squares), and ionization rate γ_3 (circles). **b**, Contour plot of photocurrent versus pulse duration and magnetic field at fixed $E_{\text{THz}} = 2.1 \times 10^4 \text{ V m}^{-1}$. The arrow indicates resonant magnetic field. Inset (left), one-dimensional colour map of linear absorption data with arrow indicating resonant magnetic field, showing a shift in resonance frequency with E_{THz} .

the energy of the $2p^+$ state, and thus requiring a higher magnetic field for resonance. However, this effect has not been calculated. Figure 4a shows dependence of the photocurrent on pulse duration for a range of magnetic fields near resonance. As the system is detuned from resonance, the frequency of the oscillations increases, and their magnitude decreases.

We modelled the dynamics of our system using the density matrix formalism for a two-level system¹⁰, modified to include relaxation into the conduction band (an ‘open’ system). The diagonal elements ρ_{11} and ρ_{22} of the density matrix ρ give the fractional occupations of the $1s$ and $2p^+$ states, respectively, while the off-diagonal elements are related to the phase coherence of the ensemble. The equations of motion of the elements ρ_{ij} of the density matrix are:

$$\frac{d\rho_{11}}{dt} = -2\Omega_R \sin(\omega_{\text{THz}}t) \text{Im}(\rho_{12}) + \gamma_1 \rho_{22}$$

$$\frac{d\rho_{22}}{dt} = 2\Omega_R \sin(\omega_{\text{THz}}t) \text{Im}(\rho_{12}) - (\gamma_1 + \gamma_3) \rho_{22}$$

$$\frac{d\rho_{12}}{dt} = i\omega_{21}\rho_{12} + i\Omega_R \sin(\omega_{\text{THz}}t)(\rho_{11} - \rho_{22}) + \gamma_2 \rho_{12}$$

Here ω_{21} is the resonant frequency of the $2p^+ - 1s$ transition, ω_{THz} the frequency of the terahertz field, $\Omega_R = eE_{\text{THz}}x_{12}/\hbar$ the Rabi frequency, e the electric charge, E_{THz} the amplitude of the terahertz electric field, and x_{12} the dipole matrix element of the $1s - 2p^+$ transition. The parameter γ_1 is the $2p^+ - 1s$ recombination rate; γ_2 is the total dephasing rate, which includes both intrinsic dephasing and inhomogeneous broadening of the transition energy; and γ_3 is the ionization rate of the $2p^+$ state. The solutions oscillate at a frequency $\Omega = [\Omega_R^2 + \delta^2 + (\text{damping terms})]^{1/2}$, where $\delta = \omega_{\text{THz}} - \omega_{21}$. On resonance, $\omega_{\text{THz}} = \omega_{21}$, and the oscillation frequency Ω is close to the Rabi frequency.

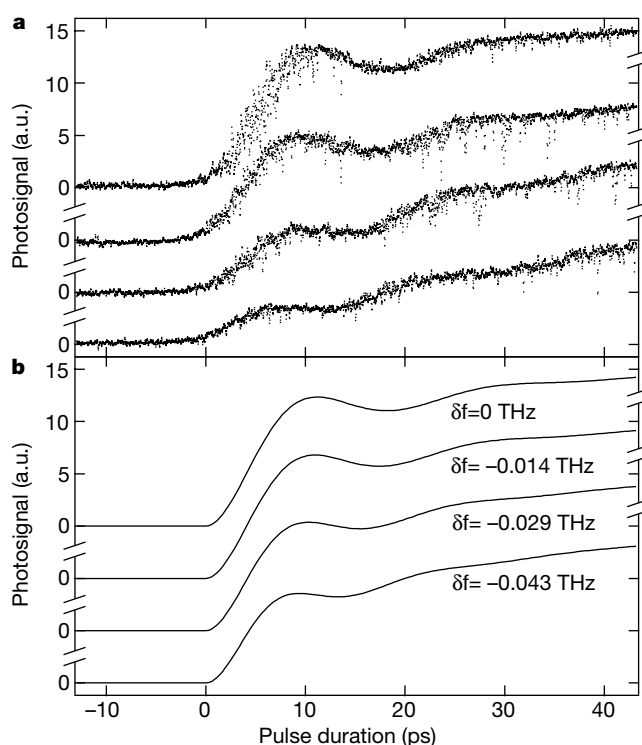


Figure 4 Photocurrent as a function of pulse duration for various detunings at a fixed terahertz field. **a**, Smoothed data measured for $E_{\text{THz}} = 2.1 \times 10^4 \text{ V m}^{-1}$, $f = 2.52 \text{ THz}$, and, beginning with top trace, magnetic field B (detuning δf) = 3.54 T (0 THz), 3.58 T (−0.014 THz), 3.61 T (−0.029 THz) and 3.64 T (−0.043 THz). **b**, Model calculations of the photocurrent as a function of magnetic field and pulse duration.

By making the rotating wave approximation¹⁰, we find equations with analytic solutions for the resonant case. The conductivity induced by a rectangular pulse of duration τ is taken to be proportional to $(1 - \rho_{11}(\tau))$, the total number of donors excited out of the ground state at the end of the terahertz pulse. The solid curves in Fig. 3a are a least-squares fit of the data to this analytic solution. The fitted curves reproduce the essential features of the data.

The inset to Fig. 3a shows the Rabi frequency and damping rates extracted from the fits as functions of E_{THz} . If we assume a dipole matrix element of 10 nm, then the Rabi frequency at a terahertz field of $3.1 \times 10^4 \text{ V m}^{-1}$ is predicted to be $4.7 \times 10^{11} \text{ rad s}^{-1}$, well within experimental error of the observed value. The fitted Rabi frequency increases roughly linearly with E_{THz} but with a non-zero intercept. The magnetic field was tuned to be resonant at the highest E_{THz} . The non-zero intercept is consistent with the detuning that results from the shift in the resonance frequency with E_{THz} (see Fig. 3b).

We calculated the curves in Fig. 4b by fixing the model parameters with a fit to the on-resonance curve, and varying only the detuning for the off-resonance curves. The frequency of the oscillations in the theoretical curves increases with increasing detuning, while the amplitude decreases, as observed in the experiment; however, we note that the amplitude decreases much more quickly with detuning in the experiment.

The mechanisms that damp the observed Rabi oscillations are extrinsic to the model qubits, and much faster than predicted intrinsic decoherence. The value of the dephasing rate γ_2 at the lowest terahertz field is, within experimental error, the same as that obtained from the linear spectra, $60 \times 10^9 \text{ s}^{-1}$. This regime is inhomogeneously broadened by a background disorder potential¹⁶. As the terahertz field is increased, the inset to Fig. 3a shows γ_2 and γ_3 increasing, consistent with a photoionization process that couples the $2p^+$ state to a higher excited state. Intrinsic decoherence of motional states of hydrogenic impurities is expected to be limited by very weak coupling to acoustic phonons^{3,22–24}. From table III of ref. 23, the contribution of acoustic-phonon coupling to the line-width (full-width at half-maximum, FWHM) of the $1s-2p$ ($m=0$) transition at zero magnetic field is predicted to be $\leq 10^{-3} \Xi^2 / (\rho a^3 v^2) \approx 0.2 \mu\text{eV}$ (where $\Xi = 8.6 \text{ eV}$ is the deformation potential, $a = 10 \text{ nm}$ is the Bohr radius, $\rho = 5,300 \text{ kg m}^{-3}$ is the density, and $v = 3,700 \text{ m s}^{-1}$ is the velocity of sound, corresponding to an intrinsic decoherence rate $\gamma_2 \leq 2 \times 10^8 \text{ s}^{-1}$. Such decoherence rates, more than 1,000 times slower than typical Rabi frequencies measured here, would enable more complex manipulations of the model qubits. Future experiments will attempt to measure the intrinsic decoherence time of $2p$ ($m=-1$) hydrogenic donor states, which are well below the continuum and hence robust to ionization by photons and phonons. □

Received 5 December; accepted 18 December 2000.

- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Sherwin, M. S., Imamoglu, A. & Montroy, T. Quantum computation with quantum dots and terahertz cavity quantum electrodynamics. *Phys. Rev. A* **60**, 3508–3514 (1999).
- Imamoglu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
- Vrijen, R. *et al.* Electron spin resonance transistors for quantum computing in silicon-germanium hetero-structures. Preprint quant-ph/9905096 at (<http://xxx.lanl.gov>) (1999).
- Bennett, C. H. & DiVincenzo, D. P. Quantum information and computation. *Nature* **404**, 247–255 (2000).
- DiVincenzo, D. P. Quantum computation. *Science* **270**, 255–261 (1995).
- Hegmann, F. A. *et al.* Time-resolved photoresponse of a gallium-doped germanium photoconductor using a variable pulse-width terahertz source. *Appl. Phys. Lett.* **76**, 262–264 (2000).
- Rabi, I. I. Space quantization in a gyrating magnetic field. *Phys. Rev.* **51**, 652–654 (1937).
- Boyd, R. W. *Nonlinear Optics* (Academic, Boston, 1992).
- Kohn, W. in *Solid State Physics* (eds Seitz, F. & Turnbull, D.) Vol. 5, 257–320 (Academic, New York, 1957).
- Yu, P. Y. & Cardona, M. *Fundamentals of Semiconductors: Physics and Materials Properties* (Springer, Berlin, 1999).
- Klaassen, T. O., Dunn, J. L. & Bates, C. A. in *Atoms and Molecules in Strong External Fields* (eds

- Schmelcher, P. & Schweizer, W.) 291–300 (Plenum, New York, 1998).
- Stillman, G. E., Wolfe, C. M. & Dimmock, J. O. Magnetospectroscopy of shallow donors in GaAs. *Solid State Commun.* **7**, 921–925 (1969).
- Burghoorn, J., Klaassen, T. O. & Wenckebach, W. T. The dynamics of shallow donor ionization in n-GaAs studied with sub-ps FIR-induced photoconductivity. *Semicond. Sci. Technol.* **9**, 30–34 (1994).
- Larsen, D. M. Inhomogeneous line broadening in donor magneto-optical spectra. *Phys. Rev. B* **8**, 535–553 (1973).
- Planken, P. C. M. *et al.* Far-infrared picosecond time-resolved measurement of the free-induction decay in GaAs:Si. *Phys. Rev. B* **51**, 9643–9647 (1995).
- Cundiff, S. T. *et al.* Rabi flopping in semiconductors. *Phys. Rev. Lett.* **73**, 1178–1181 (1994).
- Schulzgen, A. *et al.* Direct observation of excitonic Rabi oscillations in semiconductors. *Phys. Rev. Lett.* **82**, 2346–2349 (1999).
- Ramian, G. The new UCSB free-electron lasers. *Nucl. Instrum. Methods Phys. Res. A* **318**, 225–229 (1992).
- Hegmann, F. A. & Sherwin, M. S. Generation of picosecond far-infrared pulses using laser activated semiconductor reflection switches. *Proc. SPIE* **2842**, 90–105 (1996).
- Bockelmann, U. Phonon scattering between zero-dimensional electronic states: Spatial versus Landau quantization. *Phys. Rev. B* **50**, 17271–17279 (1994).
- Barrie, R. & Nishikawa, K. Phonon broadening of impurity spectral lines II. Application to silicon. *Can. J. Phys.* **41**, 1823–1835 (1963).
- Nishikawa, K. & Barrie, R. Phonon broadening of impurity spectral lines I. General theory. *Can. J. Phys.* **41**, 1135–1173 (1963).
- Stanley, C. R. *et al.* Electrical characterization of molecular beam epitaxial GaAs with peak electron mobilities up to approximately $4 \times 10^5 \text{ cm}^2/\text{V-s}$. *Appl. Phys. Lett.* **58**, 478–480 (1991).
- Stanley, C. R., Holland, M. C., Hutchins, R. H., Kean, A. H. & Johnson, N. P. in *Institute of Physics Conference Series* Vol. 112 (ed. Singer, K. E.) 67–72 (IOP Publishing, London, 1990).
- McKnight, S. W., Stewart, K. P., Drew, H. D. & Moorjani, K. Wavelength-independent anti-interference coating for the far-infrared. *Infrared Phys.* **27**, 327–333 (1987).

Acknowledgements

We thank D. K. Enyeart and C. Sean Roy for assistance with experiments, and C. J. Weinberger, D. D. Awschalom, and A. Imamoglu for critical readings of the manuscript. This work was supported by the ARO, the ONR/Medical Free-Electron Laser Program, the NSF, and Sun Microsystems.

Correspondence and requests for materials should be addressed to M.S.S. (e-mail: sherwin@physics.ucsb.edu).

Superconductivity at 39 K in magnesium diboride

Jun Nagamatsu*, Norimasa Nakagawa*, Takahiro Muranaka*, Yuji Zenitani* & Jun Akimitsu*†

* Department of Physics, Aoyama-Gakuin University, Chitosedai, Setagaya-ku, Tokyo 157-8572, Japan

† CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

In the light of the tremendous progress that has been made in raising the transition temperature of the copper oxide superconductors (for a review, see ref. 1), it is natural to wonder how high the transition temperature, T_c , can be pushed in other classes of materials. At present, the highest reported values of T_c for non-copper-oxide bulk superconductivity are 33 K in electron-doped $\text{Cs}_x\text{Rb}_{1-x}\text{C}_{60}$ (ref. 2), and 30 K in $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ (ref. 3). (Hole-doped C_{60} was recently found⁴ to be superconducting with a T_c as high as 52 K, although the nature of the experiment meant that the supercurrents were confined to the surface of the C_{60} crystal, rather than probing the bulk.) Here we report the discovery of bulk superconductivity in magnesium diboride, MgB_2 . Magnetization and resistivity measurements establish a transition temperature of 39 K, which we believe to be the highest yet determined for a non-copper-oxide bulk superconductor.

The samples were prepared from powdered magnesium (Mg; 99.9%) and powdered amorphous boron (B; 99%) in a dry box. The powders were mixed in an appropriate ratio ($\text{Mg:B} = 1:2$), ground and pressed into pellets. The pellets were heated at 973 K

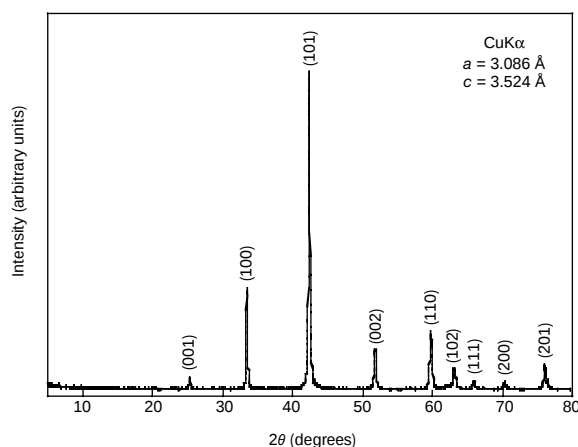


Figure 1 X-ray diffraction pattern of MgB_2 at room temperature.

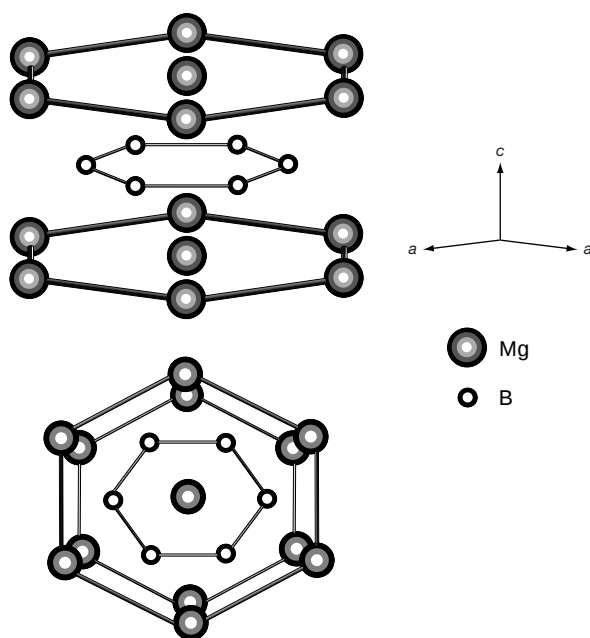


Figure 2 Crystal structure of MgB_2 .

under a high argon pressure, 196 MPa, using a hot isostatic pressing (HIP) furnace ($\text{O}_2\text{Dr.HIP}$, Kobelco) for 10 hours. Powder X-ray diffraction was performed by a conventional X-ray spectrometer with a graphite monochromator (RINT-2000, Rigaku). Intensity data were collected with $\text{CuK}\alpha$ radiation over a 2θ range from 5° to 80° at a step width of 0.02° .

Figure 1 shows a typical X-ray diffraction pattern of MgB_2 taken at room temperature. All the intense peaks can be indexed assuming an hexagonal unit cell, with $a = 3.086 \text{ \AA}$ and $c = 3.524 \text{ \AA}$. Figure 2 shows the crystal structure of MgB_2 (ref. 5), of which the space group is $P6/mmm$ (no.191). As shown in Fig. 2, the boron atoms are arranged in layers, with layers of Mg interleaved between them. The structure of each boron layer is the same as that of a layer in the graphite structure: each boron atom is here equidistant from three other boron atoms. Therefore, MgB_2 is composed of two layers of boron and magnesium along the c axis in the hexagonal lattice.

Magnetization measurements were also performed with a SQUID magnetometer (MPMSR2, Quantum Design). Figure 3 shows the magnetic susceptibility ($\chi = M/H$, where M is magnetization and H is magnetic field) of MgB_2 as a function of temperature, under

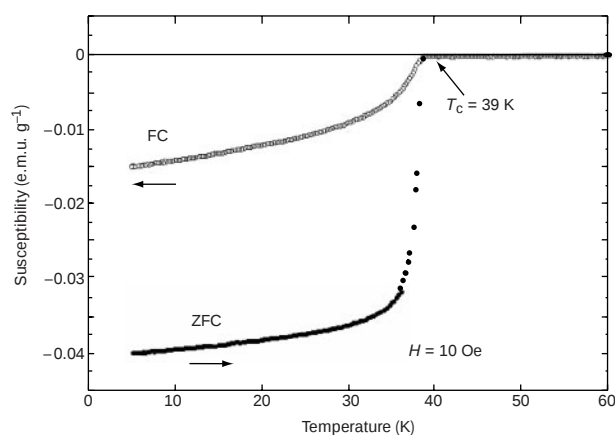


Figure 3 Magnetic susceptibility χ of MgB_2 as a function of temperature. Data are shown for measurements under conditions of zero field cooling (ZFC) and field cooling (FC) at 10 Oe.

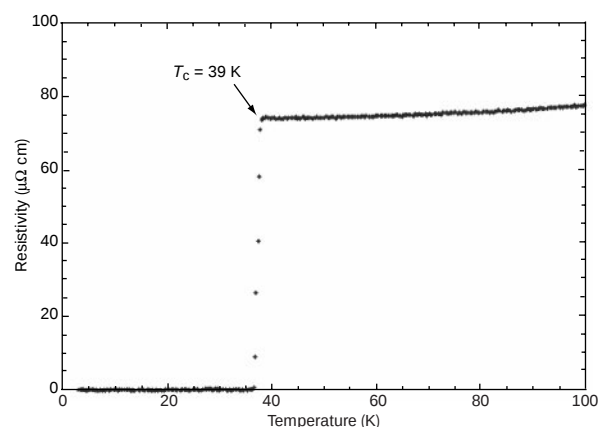


Figure 4 Temperature dependence of the resistivity of MgB_2 under zero magnetic field.

conditions of zero field cooling (ZFC) and field cooling (FC) at 10 Oe. The existence of the superconducting phase was then confirmed unambiguously by measuring the Meissner effect on cooling in a magnetic field. The onset of a well-defined Meissner effect was observed at 39 K. A superconducting volume fraction of 49% under a magnetic field of 10 Oe was obtained at 5 K, indicating that the superconductivity is bulk in nature. The standard four-probe technique was used for resistivity measurements.

Figure 4 shows the temperature dependence of the resistivity of MgB_2 under zero magnetic field. The onset and end-point transition temperatures are 39 K and 38 K, respectively, indicating that the superconductivity was truly realized in this system. $\square$

Received 24 January; accepted 5 February 2001.

1. Takagi, H. in *Proc. Int. Conf. on Materials and Mechanisms of Superconductivity, High Temperature Superconductors VI*. Physica C **341-348**, 3–7 (2000).
2. Tanigaki, K. et al. Superconductivity at 33 K in $\text{Cs}_x\text{Rb}_{1-x}\text{C}_{60}$. *Nature* **352**, 222–223 (1991).
3. Cava, R. V. et al. Superconductivity near 30 K without copper: the $\text{Ba}_{0.6}\text{K}_{0.4}\text{BiO}_3$ perovskite. *Nature* **332**, 814–816 (1988).
4. Schön, J. H., Kloc, Ch. & Batlogg, B. Superconductivity at 52 K in hole-doped C_{60} . *Nature* **408**, 549–552 (2000).
5. Jones, M. & Marsh, R. The preparation and structure of magnesium boride, MgB_2 . *J. Am. Chem. Soc.* **76**, 1434–1436 (1954).

Acknowledgements

This work was partially supported by a Grant-in-Aid for Science Research from the Ministry of Education, Science, Sports and Culture, Japan and by a grant from CREST.

Correspondence and requests for materials should be addressed to J.A. (e-mail: jun@soliton.phys.aoyama.ac.jp).

Non-thermal melting in semiconductors measured at femtosecond resolution

A. Rousse^{*}, C. Rischel[†], S. Fourmaux^{*}, I. Uschmann[‡], S. Sebban^{*}, G. Grillon^{*}, Ph. Balcou^{*}, E. Förster[‡], J.P. Geindre[§], P. Audebert[§], J.C. Gauthier[§] & D. Hulin^{*}

^{*}Laboratoire d'Optique Appliquée, ENSTA, CNRS UMR7639, École Polytechnique, Chemin de la Hunière, 91761 Palaiseau, France

[†]Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark, and Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

[‡]X-ray Optics Group, Institute of Optics and Quantum Electronics, Friedrich Schiller University Jena, Max-Wien-Platz 1, 07743 Jena, Germany

[§]Laboratoire pour l'Utilisation des Lasers Intenses, UMR7605, CNRS, École Polytechnique, CEA, Université Paris VI, École Polytechnique, 91128 Palaiseau, France

Ultrafast time-resolved optical spectroscopy has revealed new classes of physical¹, chemical² and biological³ reactions, in which directed, deterministic motions of atoms have a key role. This contrasts with the random, diffusive motion of atoms across activation barriers that typically determines kinetic rates on slower timescales. An example of these new processes is the ultrafast melting of semiconductors, which is believed to arise from a strong modification of the inter-atomic forces owing to laser-induced promotion of a large fraction (10% or more) of the valence electrons to the conduction band^{1,4–12}. The atoms immediately begin to move and rapidly gain sufficient kinetic energy to induce melting—much faster than the several picoseconds required to convert the electronic energy into thermal motions¹³. Here we present measurements of the characteristic melting time of InSb with a recently developed technique of ultrafast time-resolved X-ray diffraction^{14–19} that, in contrast to optical spectroscopy, provides a direct probe of the changing atomic structure. The data establish unambiguously a loss of long-range order up to 900 Å inside the crystal, with time constants as short as 350 femtoseconds. This ability to obtain the quantitative structural characterization of non-thermal processes should find widespread application in the study of ultrafast dynamics in other physical, chemical and biological systems.

Optical measurements on semiconductors have confirmed that an intense femtosecond laser pulse instantly produces a dense electron–hole plasma^{1,5–7,9}. However, the signature of the expected liquid state is difficult or impossible to separate from contributions from purely electronic effects^{7–9}. We have used a laser-produced plasma X-ray source to study the femtosecond atomic response of the InSb semiconductor to strong 120-fs laser pulse excitation by ultrafast time-resolved X-ray diffraction. The characterization of the excited electron–hole plasma has been obtained through time-dependent reflectivity measurements performed with the same equipment as X-ray acquisition (see Methods for details). The X-ray probing depths range from 3,500 to 540 Å, the latter being much smaller than the optical excitation depth (1,000 Å).

The transient integrated intensity of the {111} Bragg reflection is shown in Fig. 1 for three different probing depths: 3,500, 1,700 and 1,200 Å, and at exciting fluences of ~120 mJ cm⁻² (130, 119 and 110 mJ cm⁻²). The intensity drops by 22%, 40% and 60% in 350, 420 and 610 fs as estimated from the exponential curves represented by the solid lines. Thereafter the signal remains constant for several picoseconds. On this timescale we observed neither shifts nor variations of the width of the rocking curve with the present resolution (20 arcsec). A shift of the Bragg reflection would have

indicated that the {111} atomic plane spacing had been reduced or increased, whereas a broadening would have indicated that the spacing had become inhomogeneous within the depth of probing. The integrated intensity reflects the order of the lattice: it is highest when the diffracting atoms are confined to narrow planes and decreases if the distribution of atoms is broadened (at least when the X-ray penetration depth is determined by absorption rather than by extinction by diffraction, as occurs for the diffraction of Si K_α on InSb). The observed subpicosecond intensity drop is a definite signature of high ultrafast atomic disordering.

What is the nature of the disordering? Ordinary melting is nucleated at the surface (or at intrinsic defects) and occurs when the forces that stabilize the lattice, mediated by the electrons, are no longer sufficient to counterbalance the thermal motion of the atoms. Then the liquid front²⁰ moves by at most the speed of sound (5 × 10³ m s⁻¹ for InSb), 40-fold slower than the highest value retrieved from the present data (2 × 10⁵ m s⁻¹). In addition, transfer of energy from excited electrons to thermal motion of the lattice is expected to take many picoseconds¹³, so thermal processes cannot account for our data. The transfer of energy to the longitudinal optical (LO) phonons is much faster. First, the sudden change of interatomic potential induced by the laser pulse immediately excites optical phonons. Furthermore, the LO phonon emission time from excited electrons is ~200 fs (ref. 21) if screening is neglected, so a high density of electrons can rapidly energize these vibrations appreciably, although it takes several picoseconds to transfer all the energy^{9,17,22}. The oscillations of the atoms could give a decrease in the diffracted energy, by analogy with the (thermal) Debye–Waller effect (see Supplementary Information). However, the Debye–Waller effect gives a decrease in diffracted intensity of only 7% close to the melting temperature (800 K)²³, so the decreases of more than 50% that we observe would require displacements far larger than the critical amplitude for thermal melting. We believe that the non-thermal formation of a liquid film on the sample surface is the most plausible explanation of our observations.

Analysis of the sample long after the excitation revealed the formation of craters with a thickness of ~200 Å, which is smaller

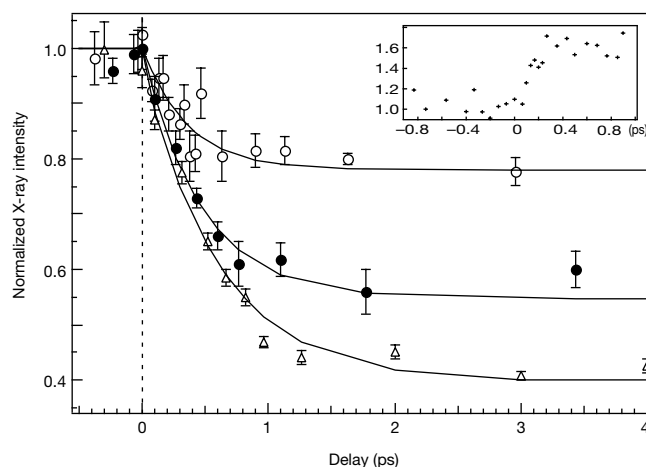


Figure 1 Normalized X-ray diffracted intensity for three different X-ray probing thicknesses as a function of the delay after the exciting laser pulse set at an absorbed laser fluence of ~120 mJ cm⁻² (130 mJ cm⁻² for 3,500 Å (open circles), 119 mJ cm⁻² for 1,700 Å (filled circles) and 110 mJ cm⁻² for 1,200 Å (open triangles)). The timescales of the measured duration of the phase transition were 350 fs for 3,500 Å, 420 fs for 1,700 Å and 610 fs for 1,200 Å. The solid lines are exponential fits of the experimental results and the error bars represent ±1 s.d. At negative delays, the X-rays probe the unperturbed InSb crystal. The inset shows the time-resolved reflectivity at 610 nm for the sample with a probe depth of 3,500 Å, obtained in exactly the same set-up as the X-ray data (angle of incidence of 80°).

than the depth over which the lattice is disordered. This rules out the possibility of a solid–plasma transition and further supports the notion that disordering at the earlier stage was high enough to produce melting, followed by ablation of a superheated liquid²⁴ and recrystallization. A full recovery of the X-ray-diffracted signal in the centre of the excited area is observed to occur within a few nanoseconds. Optical measurements show that a solid–plasma transition occurs only for intensities above 200 mJ cm^{-2} .

In the simplified case that the molten layer lies on top of an unperturbed solid, the integrated diffracted intensity is $A = A_0 \exp(-d/l)$, where A_0 is the intensity observed without excitation, d is the thickness of the molten layer and l is the probe depth (defined in Methods). In Fig. 2 we have plotted d derived from experiments with excitation fluences from 3 to 130 mJ cm^{-2} and with four different probe depths, together with the observed duration of the melting process. The solid line fits a simple model, which assumes that melting occurs in a depth z if the intensity at this point exceeds a critical value I_c . The attenuation of light is taken to occur by one-photon and two-photon absorption with coefficients α and β : $dI/dz = -\alpha I - \beta I^2/\tau$, where τ is the pulse duration (120 fs). If the intensity at the surface is I_0 , the thickness of the molten layer is

$$d = \frac{1}{\alpha} \ln \left[(1 + \alpha/\beta I_c) \frac{I_0}{I_0 + \alpha/\beta} \right]$$

The fit obtained with β and I_c as free parameters (α is $1.0 \times 10^5 \text{ cm}^{-1}$) (ref. 25) is seen to be satisfactory. We obtain $I_c = 9.5 \pm 1.3 \text{ mJ cm}^{-2}$ and $\beta = 780 \pm 190 \text{ cm GW}^{-1}$. This value of I_c is approximately equal to the threshold for irreversible surface damage determined from post-mortem analysis of samples (ref. 9, and A.R. and S.F., unpublished observations).

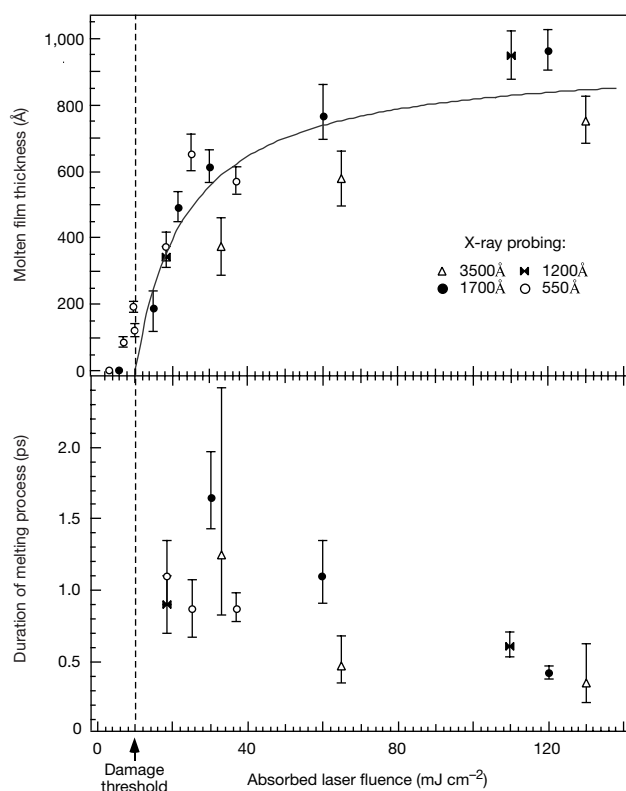


Figure 2 Thickness of the molten layer and duration of the melting process as functions of the absorbed laser fluence (corrected for reflectivity). The solid line is a fit to the one-photon and two-photon model described in the text. Only points above 10 mJ cm^{-2} were used in the fitting procedure. Below 15 mJ cm^{-2} most time constants were $>3 \text{ ps}$ and reproducibility was poor (data not shown).

Our observations are consistent with previous optical measurements on Si, GaAs and InSb^{1,5,7,9}. In these experiments, the reflectivity reaches a plateau (the reflectivity of a liquid) in less than 1 ps for fluences exceeding several times the melting threshold. At lower fluences, the continuous increase in reflectivity continues for a few picoseconds. To connect with previous work, we have measured the time-dependent reflectivity at 610 nm. The inset in Fig. 1 shows the results for the sample with an X-ray probe depth of 3,500 Å. The reflectivity rapidly increases 1.6-fold, which is in good agreement with the data of ref. 9 when the different angles of incidence are taken into account. For Si and GaAs, non-thermal melting seems to occur only above ~ 1.5 -fold the melting threshold⁷. Just above threshold, melting takes tens of picoseconds, which is consistent with the classical picture of thermal melting. In agreement with this, we measure time constants of several picoseconds at excitation fluences below 15 mJ cm^{-2} . However, the values vary considerably between different experimental runs, probably owing to a very high sensitivity to the fluence. This makes it difficult to determine the exact threshold for non-thermal melting.

A crude estimate of the density N_e of excited electrons at the surface can be made by using the following expression⁷: $N_e = I(\alpha + \beta I/\tau)/h\nu$, in which $h\nu$ is the photon energy (1.55 eV). With $\alpha = 10^5 \text{ cm}^{-1}$ and $\beta = 780 \text{ cm GW}^{-1}$, a fluence of 10 mJ cm^{-2} gives an electron density of $7 \times 10^{21} \text{ cm}^{-3}$. This is quite close to the threshold density for ultrafast melting of $\sim 10^{22} \text{ cm}^{-3}$ found in calculations and experiments on Si and GaAs, in particular considering that the total valence electron density of InSb is $\sim 70\%$ that of GaAs. The actual threshold for InSb might also be slightly higher, as we have just discussed.

The model of lattice destabilization after laser-induced high carrier concentration^{10–12} predicts high disordering (displacements up to 1 Å) in silicon within 120 fs of the excitation pulse when 15% of the valence electrons are driven to the conduction band. For other semiconductors with larger atomic mass M and bond length a , the time for the transition has been predicted to be inversely proportional to the phonon frequencies¹⁰, which scale as $\omega \propto d^{-2} M^{-1/2}$. If we ignore the differences in band structure, the time of the transition in InSb is then expected to be 350 fs, corresponding to the response that we observe at an exciting fluence of 120 mJ cm^{-2} . However, the duration of the melting process is significantly longer at low intensities. This slowing down was previously observed qualitatively in optical measurements^{7,9}, but a quantitative determination of the melting time of InSb was not attempted. It should be noted that the melting times in Fig. 2 are plotted as a function of the fluence absorbed at the surface. The X-rays also probe deeper regions, which experience a lower intensity and probably melt more slowly than the surface layers. The recovered time constant is an average over the entire molten region. As discussed above, thermal melting of deeper layers cannot contribute a significant component because the melt front in this case advances by at most 50 Å in 1 ps.

The measured time constants of more than 1 ps seem to challenge the ‘barrier-free’ view of non-thermal melting. If the atoms really could move freely after excitation, they would do so at least with their initial thermal velocities at 300 K. The diffracted intensity decreases by a factor 1/e when the atoms have moved an average distance $a/2\pi$, where a is the lattice spacing¹⁴, and the time required to do this would be $a/[2\pi\sqrt{(kT/m)}] = 400 \text{ fs}$. Because it is here assumed that no additional energy is acquired from the hot electrons, this is a quite conservative (high) estimate, but our data show melting to occur several times more slowly. At high fluences the melting time in fact decreases to $\sim 400 \text{ fs}$, but this is probably fortuitous because the strong destabilization of the lattice must give the atoms additional energy. One could imagine that the time constant was determined by diffusion of hot electrons and holes, such that only a thin layer at the surface was initially highly excited,

diffusion of hot carriers subsequently bringing the density in deeper layers up to the threshold value for melting. We have solved the diffusion equation numerically with the initial density of carriers falling off exponentially from the surface: the broadening turns out to be rather modest. As an example, diffusion expands the molten layer by only 40% when the initial density at the surface is fourfold the threshold density. Furthermore, the relative expansion of the molten region by diffusion is larger at higher surface excitations, and takes a longer time. In reality, we do not observe substantial picosecond components in the response at high fluence, so diffusion does not seem to have a major role on this timescale.

We conclude that the response time reflects the intrinsic time for ultrafast melting. The atoms must therefore—at intermediate excitation fluences—perform many complete LO vibrations (with a period of 160 fs) before going into the liquid phase, which suggests the existence of an energy barrier to be surmounted. Because the phonons are far from thermal equilibrium, we can categorize ultrafast melting of InSb as a non-thermally activated process. The time required for passage of the barrier depends on its height and on the average energy in the oscillations. The faster melting at intense excitation can be explained in part by the lowering of the barrier suggested by theory¹², and in part by stronger flow of energy into the vibrations. We believe that this picture of ultrafast melting is in agreement with all previous experimental work. We have not attempted quantitative estimates because the phonon dynamics at high free-carrier density is complex: screening tends to decrease the emission time²², whereas an instantaneous creation of phonons stimulates further emission owing to boson effects. Accurate theoretical prediction of the duration of the process as a function of fluence is very demanding because of the difficulty in performing the numerical simulation with sufficient atoms to retrieve a long-range evolution of the matter. Our results might serve as benchmarks for attempts to establish a quantitative theory. Below the damage threshold the energy transferred to the optical phonons is insufficient to produce melting in a few picoseconds. The small signal we observe at these excitation fluences might be due to a non-thermal Debye–Waller effect.

The thermal response of the lattice, revealed by a shift of the rocking curve (compression), appears from ~4 ps at fluences above the damage threshold. The shift then gradually increases and at the same time the rocking curve broadens. The magnitude of the shift indicates a maximum strain as high as -0.3%. Later, a component with the opposite shift appears, indicating an expanded lattice. If the propagation of the heat-induced strain is described by a wave equation (acoustic), one can obtain a simple analytical solution²⁶, which consists of a static expansion near the surface and a strain wave travelling into the solid at the speed of sound, with a compressive front and a rarefaction tail as observed in a recent picosecond time-resolved X-ray diffraction experiment¹⁵. X-ray diffraction simulations with this strain profile show that the rocking curves observed at long delays are consistent with an intense compressive wave travelling at the speed of sound ($5 \times 10^3 \text{ m s}^{-1}$).

The present data answer a standing question in solid-state physics by giving an unambiguous, quantitative characterization of ultrafast melting of semiconductors. The possibility of direct structural measurements of non-thermal processes will also influence many other domains of science. In ultrafast science, one can resolve the very few picoseconds that are the fastest times for vibrational relaxation in solids and biological macromolecules, then extend into the subpicosecond realm, where the atomic motions are unaffected by relaxation processes. Rather than just heating, excitation on this timescale induces a directed motion of the atoms, such that entirely new phenomena might occur^{2,3,10,11}. For example, it was shown recently that the ultrafast dissociation of CO from the protein cytochrome *c* oxidase is governed by coherent motions²⁷. Our results yield the first structural measurement of a non-thermal process in real time, with the method of ultrafast X-ray

diffraction. As this technique matures^{14–17}, we can expect further insights into such processes in physics, chemistry and structural biology. □

Methods

The experimental set-up was essentially as described previously¹⁴, with the possibility of adding an optical probe. In brief, the sample was excited by a laser delivering pulses of 800 nm and 120 fs, and the response was probed by X-ray radiation at 7.13 Å generated by focusing a laser pulse of 23 mJ, 800 nm and 120 fs from a second arm of the visible laser on a silicon target. Under the carefully controlled conditions, the monochromatic K_α X-ray radiation is expected to retain a duration similar to that of the initial laser pulse. The emitted X-rays were collected over a solid angle of 7.4×10^{-3} steradians (sr) by a toroidally bent quartz (100) crystal and focused on the InSb sample. Because the width of the quartz rocking curve was much smaller than that of InSb, the observed diffraction patterns depended only on the structure of the probed sample. The $100 \mu\text{m} \times 100 \mu\text{m}$ X-ray focal spot contained 5×10^4 photons per shot. The first-order Bragg reflection of InSb in the (111) orientation (Bragg angle 72.3°) was detected by a cooled X-ray charge-coupled device (CCD) camera, with the estimated number of diffracted photons reaching ~500 in one laser shot. Both (111), (100) and asymmetrically cut (100) InSb crystals were used to change the probe depth l of the X-ray beam by varying the angle of incidence θ with respect to the surface. Because the angle between incoming and outgoing rays in all cases was 35.4° , the probe depth through a disordered layer on the surface was the penetration length $7,100 \text{ Å}$ divided by $1/\sin \theta + 1/\sin(\theta + 35.4^\circ) = \sin^{-1}(\theta) + \sin^{-1}(\theta + 35.4^\circ)$. The diffracted intensity was normalized for each shot by using the emitted intensity detected by a photodiode with a Be window facing the Si target. The optical probe beam was a continuum produced by focusing 20 μJ of the infrared laser pulse into a water cell. The reflected p-polarized signal was collected with a cooled 16-bit CCD camera through a spectrometer. The intensity was kept well below 0.005 J cm^{-2} to avoid any excitation of the sample. The $400\text{-}\mu\text{m}$ focal spot diameter of the exciting pulse was set more than threefold larger than the focal spots of the probing beams so as to probe a homogeneously excited area. An experimental run consisted of three probe shots without excitation, one shot with both exciting and probe beam and three more probe shots without excitation to probe the expected damaged area. To obtain one data point for a given pump-probe delay this sequence was repeated six times, with a fresh surface being exposed each time.

The threshold intensity for the solid–plasma transition was measured with the method of Pronko *et al.*²⁸. Post-mortem analysis of the samples was performed with optical and atomic force microscopy, with optical and stylus profilers. X-ray diffraction simulations were performed with SpotX software.

Received 7 June; accepted 12 December 2000.

- Shank, C. V., Yen, R. & Hirlimann, C. Femtosecond time-resolved surface structural dynamics of optically excited silicon. *Phys. Rev. Lett.* **51**, 900–902 (1983).
- Zewail, A. H. Femtochemistry: recent progress in studies of dynamics and control of reactions and their transition states. *J. Phys. Chem.* **100**, 12701–12724 (1996).
- Vos, M. H. & Martin, J.-L. Femtosecond processes in proteins. *Biochim. Biophys. Acta Bioenerget.* **1411**, 1–20 (1999).
- Wood, R. F., White, C. W. & Young, R. T. (eds) in *Pulsed Laser Processing of Semiconductors* (Academic, Orlando, 1984).
- Shank, C. V., Yen, R. & Hirlimann, C. Time-resolved reflectivity of femtosecond-optical-pulse-induced phase transitions in silicon. *Phys. Rev. Lett.* **50**, 454–4520 (1983).
- Huang, L., Callan, J. P., Glezer, E. N. & Mazur, E. GaAs under intense ultrafast excitation: Response of the dielectric function. *Phys. Rev. Lett.* **80**, 185–188 (1998).
- Sokolowski-Tinten, K., Bialkowski, J. & von der Linde, D. Ultrafast laser-induced order–disorder transitions in semiconductors. *Phys. Rev. B* **51**, 14186–14198 (1995).
- Sokolowski-Tinten, K., Schulz, H., Bialkowski, J. & von der Linde, D. Two distinct transitions in ultrafast solid-liquid phase transformations of GaAs. *Appl. Phys. A* **53**, 227–234 (1991).
- Shumay, I. L. & Höfer, U. Phase transformations of an InSb surface induced by strong femtosecond laser pulses. *Phys. Rev. B* **53**, 15878–15884 (1996).
- Stampfli, P. & Bennemann, K. H. Theory for the laser-induced femtosecond phase transition of silicon and GaAs. *Appl. Phys. A* **60**, 191–196 (1995).
- Silvestrelly, P. L., Alavi, A., Parrinello, M. & Frenkel, D. Ab initio molecular dynamics simulation of laser melting of silicon. *Phys. Rev. Lett.* **77**, 3149–3152 (1996).
- Graves, J. S. & Allen, R. E. Response of GaAs to fast intense laser pulse. *Phys. Rev. B* **58**, 13627–13633 (1999).
- Laude, L. D. *Cohesive Properties of Semiconductors Under Laser Irradiation*. (NATO ASI Series Martinus Nijhoff Vol. 69, The Hague, 1983).
- Rischel, C. *et al.* Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390**, 490–492 (1997).
- Rose-Petruck, C. *et al.* Picosecond-milllingström lattice dynamics measured by ultrafast X-ray diffraction. *Nature* **398**, 310–312 (1999).
- Larsson, J. *et al.* Ultrafast structural changes measured by time-resolved X-ray diffraction. *Appl. Phys. A* **66**, 587–591 (1998).
- Chin, A. H. *et al.* Ultrafast structural dynamics in InSb probed by time-resolved X-ray diffraction. *Phys. Rev. Lett.* **83**, 336–339 (1999).
- Siders, C. W. *et al.* Detection of non-thermal melting by ultrafast x-ray diffraction. *Science* **286**, 1340–1342 (1999).
- Lindenberg, A. M. *et al.* Time-resolved X-ray diffraction from coherent phonons during a laser-induced phase transition. *Phys. Rev. Lett.* **84**, 111–114 (2000).
- Phillpot, S. R., Yip, S. & Wolf, D. How do crystals melt? *Comput. Phys.* **3**, 20–31 (1989).
- Hellwege, K. H. (ed.) in *Landolt-Börnstein Semiconductors Vol. 17-a Group IV elements and III-V Compounds* (Springer, Berlin, 1991).
- Yoffa, E. J. Dynamics of dense laser-induced plasmas. *Phys. Rev. B* **21**, 2415–2425 (1980).
- Vetelino, J. F., Gaur, S. P. & Mitra, S. S. Debye–Waller factor for zinc-blende-type crystals. *Phys. Rev. B*

- 5, 2360–2366 (1971).
24. Sokolowski-Tinten, K. *et al.* Transient states of matter during short laser pulse ablation. *Phys. Rev. Lett.* **81**, 224–227 (1998).
25. Aspnes, D. E. & Studna, A. A. Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6 eV. *Phys. Rev. B* **27**, 985–1008 (1983).
26. Thomsen, C., Grahn, H. T., Maris, H. J. & Tauc, J. Surface generation and detection of phonons by picosecond light pulses. *Phys. Rev. B* **34**, 4129–4137 (1986).
27. Liebl, U. *et al.* Coherent reaction dynamics in a bacterial cytochrome *c* oxidase. *Nature* **401**, 181–184 (1999).
28. Pronko, P. P., Dutta, S. K., Du, D. & Singh, R. K. Thermophysical effects in laser processing of materials with picosecond and femtosecond pulses. *J. Appl. Phys.* **78**, 6233–6240 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank G. Hamoniaux, P. Rousseau, K. Ta Phuoc and A. Alexandrou from the Laboratoire d'Optique Appliquée, and T. Moreno from Caminotec Inc, for support; and Veeco Instruments Inc. and A. Semerok from CEA-Saclay for the use of the atomic force microscope (AFM) and the profilers. This work was supported by the European Community.

Correspondence should be addressed to A.R. (e-mail: rousse@ensta.fr).

Propagating solitary waves along a rapidly moving crack front

Erhan Sharon, Gil Cohen & Jay Fineberg

The Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel

A rapidly moving crack in a brittle material is often idealized¹ as a one-dimensional object with a singular tip, moving through a two-dimensional material. However, in real three-dimensional materials, tensile cracks form a planar surface whose edge is a rapidly moving one-dimensional singular front. The dynamics of these fronts under repetitive interaction^{2–4} with material inhomogeneities (asperities) and the morphology^{5–11} of the fracture surface that they create are not yet understood. Here we show that perturbations¹² to a crack front in a brittle material result in long-lived and highly localized waves, which we call 'front waves'. These waves exhibit a unique characteristic shape and propagate along the crack front at approximately^{13–15} the Rayleigh wave speed (the speed of sound along a free surface). Following interaction, counter-propagating front waves retain both their shape and amplitude. They create characteristic traces along the fracture surface, providing cracks with both inertia and a new mode of dissipation. Front waves are intrinsically three-dimensional, and cannot exist in conventional two-dimensional theories of fracture¹. Because front waves can transport and distribute asperity-induced energy fluctuations throughout the crack front, they may help to explain how cracks remain a single coherent entity, despite repeated interactions with randomly dispersed asperities.

Much recent research has focused on crack front coherence and roughening. Simplified models (mode III) of fracture^{2–4} as well as more general models of moving fronts⁹ propagating through heterogeneous media have predicted intrinsic roughness of crack fronts. Recent theoretical studies of fracture^{13–15} under tension (mode I), however, have predicted that the interaction of a crack front with localized material inhomogeneities (either 'hard' or 'soft' spots in the path of the crack) could lead to a new type of elastic wave that propagates along the crack front. Analysis¹³ of perturbations to a straight crack front¹⁶ predicted that waves, describing local velocity variations within the fracture plane would propagate,

relative to the medium, at velocities slightly below the Rayleigh wave speed, V_R . The stability of these waves depends on the velocity dependence of the fracture energy, defined as the energy needed to create a unit area of fracture surface. Increase (decrease) of the fracture energy with the crack velocity causes these linear waves to decay (grow).

In recent three-dimensional computations¹⁴ of tensile fracture in an elastic material, asperities (localized material inhomogeneities) along a crack's path were numerically shown to generate persistent, localized pulses of in-plane velocity fluctuations that propagated (as in ref. 13) at velocities slightly lower than V_R . The pulses initially decayed with a decay length of the order of the asperity diameter. They then continued to propagate with constant shape and amplitude. Here, for the first time, to our knowledge, this type of wave is observed. The experimentally observed waves are, in some respects, similar to the predicted ones. They possess, however, two important and unexpected attributes; a unique characteristic shape and existence both within and normal to the fracture plane.

Our experiments were performed on brittle soda-lime glass plates. In this material the fracture energy, as assumed in ref. 14, is nearly independent¹⁷ of the crack velocity. In Fig. 1 we present a typical photograph and surface profile measurement of fracture surfaces formed by crack fronts meeting asperities that were introduced (see Methods) at the plate surface. At the point of interaction, well-defined tracks are generated that run along the fracture surface. The tracks are created by a new type of localized wave, front waves, which propagate along the crack front. In contrast to the in-plane waves predicted in refs 13 and 14, the out-of-plane surface structure shows that these waves are three-dimensional entities. Front waves can be generated by either externally placed asperities or, beyond the critical velocity for the microbranching instability¹², spontaneously, by localized microscopically branching events, which effectively create a local perturbation of the fracture energy. Front waves arriving at sample edges are reflected (Fig. 1) with insignificant loss of amplitude. Both the long lifetime and the highly localized nature of these pulse-like waves are evident. Front waves form mirror images on opposing crack faces. Thus the tracks are formed by deflection of the crack front normal to the fracture plane.

Upon meeting, counter-propagating front waves retain both their form and amplitude (see Fig. 1a). Front-wave intersections, as presented in Fig. 2, provide measurements of both front-wave velocity and the local orientation of the crack front. Two counter-propagating pulses moving with velocity V_{FW} relative to the medium, along a front propagating at velocity v , will intersect at an angle α given by $\cos(\alpha/2) = v/V_{FW}$, when the crack front is

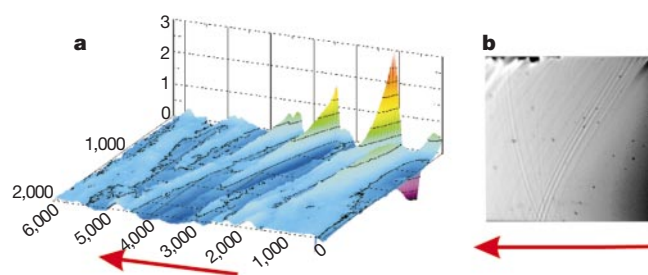


Figure 1 Highly localized propagating front waves form tracks along the fracture surface. **a**, Profilometer scans (with units in μm) and **b**, photographs of the fracture surface showing tracks formed by front waves initiated by the interaction of a crack front moving at $0.4V_R$ with an externally imposed asperity at the sample surface. The highly localized waves initially undergo rapid decay as they propagate along the crack front. They then stabilize with negligible decay and dispersion as they continue to propagate. Nearly lossless reflection (**a**, right side and **b**, centre) at the sample surfaces is typically observed. The 3-mm arrows indicate the front's direction of propagation.

normal to the direction of propagation. If the front is oriented at an angle β relative to $\mathbf{v}$, then $v/V_{FW} = \cos(\alpha/2)\cos(\beta)$.

Independent measurements¹² of v , together with direct measurements of α and β provide us with a precise measurement of V_{FW} as shown in Fig. 2c. The measurements, obtained for a wide range of v , all yield values of V_{FW} near V_R with a mean value of $(1.0 \pm 0.05)V_R$. The values of V_{FW} were derived using locally averaged values of v . The scatter in the figure results from the fluctuations of the instantaneous value of v (see Fig. 2d), owing both to the micro-branching instability¹² and to the pulses themselves. The value of V_{FW} agrees well with predictions^{13,14} for the velocity of in-plane pulses. Knowing V_{FW} measurements of α and β provide us with a new and useful tool to determine both the instantaneous value of v , at distinct points throughout the front, and the local orientation of the front. This is shown in Fig. 2d, where v , obtained in this way, agrees well with independent measurements performed at the outer edges of the fractured plates. This tool may well prove useful in the analysis of fracture surfaces. For example, analysis of characteristic front-wave traces on fractured rock faces^{10,11} will, by enabling measurement of the velocity of the crack that formed them, provide valuable information about geophysical processes involving brittle fracture.

Pulse amplitudes (see Fig. 1) initially undergo rapid decay. As Fig. 3 shows, front waves appear over a wide range of length scales. Defining a , the pulse width, as the distance between the first minimum and first maximum of the fracture surface amplitudes generated by an asperity, we see (Fig. 3a) that front-wave decay scales linearly with a . Figure 3b shows that this decay is exponential as a function of the propagation distance with a characteristic decay length of $(0.9 \pm 0.1)a$. After traversing a distance of approximately $2a$, front waves stabilize and continue to propagate with little loss of amplitude. Pulses undergo nearly lossless reflections at sample edges and, as shown in Fig. 3c, pulses may persist for multiple reflections.

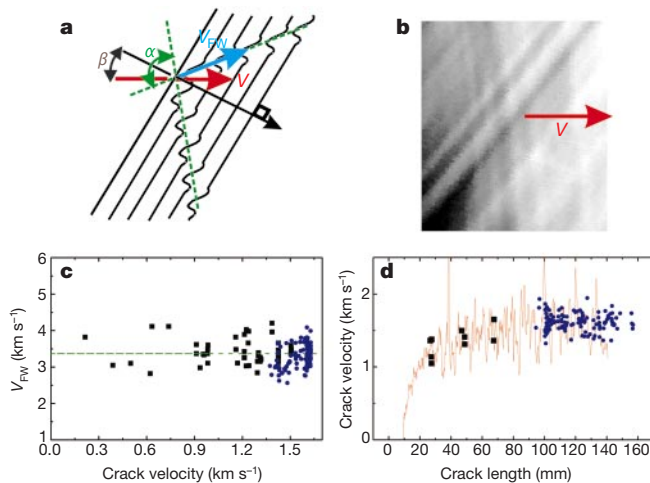


Figure 2 Determination of front-wave velocities by analysis of their tracks on the fracture surface. **a**, A schematic representation of the motion of counter-propagating front waves along a tilted crack front moving at velocity v and oriented at an angle β relative to the propagation direction. From the angle α , defined by intersecting tracks along the fracture surface, the velocity, V_{FW} , of the waves can be determined. **b**, A typical photograph showing a number of intersecting tracks along a fracture surface. **c**, Using $V_{FW} = v\cos(\beta)/\cos(\alpha/2)$, V_{FW} is calculated, where v is obtained from independent velocity measurements at the sample surfaces. The results show that V_{FW} is constant with v and equal to V_R (indicated by the dashed line). **d**, v can now be determined using $V_{FW} = V_R$. Shown is a comparison between the measured velocities (line) at the sample surfaces with the velocities calculated away from the surfaces obtained using intersecting front waves. In **c** and **d** the squares (circles) describe measurements where the front waves were initiated by externally imposed asperities (spontaneously nucleating micro-branches¹²).

When reflections from the sample edge occur, strong fluctuations corresponding to their arrival are detected in the crack velocities measured at the points of reflection. These velocity fluctuations are, in fact, the in-plane component of the front wave. As changes¹ in v can only be caused by local changes in the amount of energy flowing into a crack's tip, the fluctuations in v caused by front waves provide direct evidence that these pulses momentarily transmit energy to the crack tip.

Localized tracks have long been observed on fracture surfaces formed in brittle materials. These tracks, called Wallner lines^{18,19}, have traditionally been interpreted as being formed by the interaction of radially propagating shear waves, generated by a local source (such as an asperity) and a crack front. There are, however, essential differences, highlighted by Figs 2 and 3, between the front waves described above and surface markings attributed to shear-wave interaction. Whereas front waves first decay exponentially and then propagate with constant amplitude over long distances, shear waves, generated by a point asperity, decay radially. Thus, markings formed by the latter mechanism would rapidly and continuously decay in amplitude. The mean value of V_{FW} (Fig. 2) is also significantly less ($\sim 2\sigma$) than the shear-wave velocity.

Surprisingly, front waves in steady state motion have a unique scale-independent shape. Initial front-wave profiles correspond to asperity shapes (Fig. 4a) and are quite varied. Within a single decay length, however, the initial wave packet narrows and evolves to the characteristic profile presented in Fig. 4b. This unique shape is independent of the front wave's spatial scale (which spanned over an order of magnitude in our experiments), and amplitude. This asymptotic shape (when scaled as in Fig. 4) is attained well before the front-wave decay phase concludes. Like its decay length, the

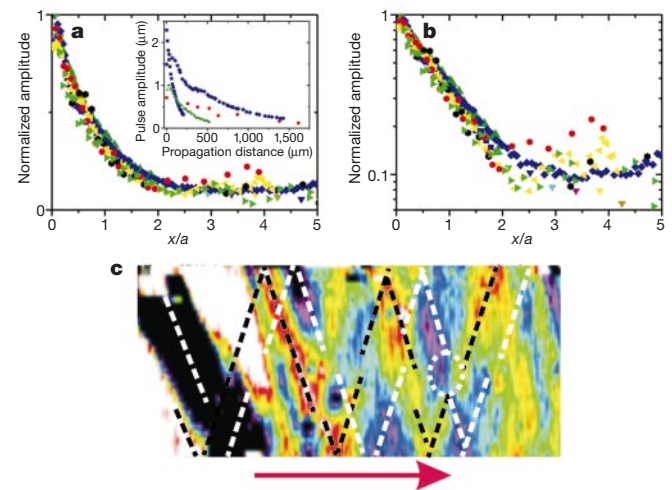


Figure 3 Front waves undergo a rapid initial decay. They then stabilize and propagate with little loss of amplitude. **a**, The initial decay profiles collapse onto a single curve, when scaled by their width a . Shown are data from seven different pulses (the unscaled decays of four of these are presented in the inset) where $a = 0.92$ mm (circles), 0.11 mm (squares), 0.26 mm (up-triangles), 0.87 mm (right triangles), 0.45 mm (left triangles), 0.65 mm (diamonds) and 0.99 mm (hexagons). **b**, The initial decays (symbols as in **a**) are exponential with a characteristic decay length of $(0.9 \pm 0.1)a$. In both **a** and **b** the front waves stabilize after propagating a distance of $\sim 2a$. The waves then continue to propagate at a nearly constant amplitude. **c**, The height of the fracture surface (scale and colour key as in Fig. 1) as a function of position, where the full width of the sample is shown. The dotted lines (black for peaks, white for depressions) show the progression of two front waves along the fracture surface. Upon reaching steady state, the waves shown undergo over six reflections at the sample surfaces. We note both the constructive and destructive interference of the waves as they pass through each other. Upon interaction, the waves retain their form but are slightly phase-shifted (indicated by breaks in the dotted lines). One such shift is highlighted (dotted ellipse). The 3-mm arrow shows the propagation direction of the front.

scale of a front wave is determined by the width of the initial perturbation, a (Fig. 4b inset). At the conclusion of their decay phase, both the width and amplitude of front waves scale roughly with a . Thus, although the size of a in the laboratory is limited ($100\text{ }\mu\text{m} < a < 1,500\text{ }\mu\text{m}$), front waves should be visible at large scales (for example, in geophysical phenomena).

Their characteristic profile and the distinctive lack of pulse dispersion are evidence that front-wave states are intrinsically nonlinear entities. If front waves were simply linear wavepackets, their shape would be highly dependent on the form of the initial conditions, scale-dependent attenuation and linear dispersion. For example, any degree of linear dispersion would generate a continuously evolving pulse profile. The observation that a well defined profile is obtained for a variety of spatial scales therefore suggests that nonlinear focusing, perhaps analogous to processes that occur in classical soliton formation, are at play. (Classical solitons²² are localized nonlinear waves formed by the balance between dispersion and nonlinear growth.) The coupling between in-plane and out-of-plane motion in front waves is further evidence of their nonlinear character, as recent calculations^{20,21} show that no such coupling, to linear order in the perturbation, occurs.

Additional evidence of the nonlinear character of front waves is provided by their interactions. When counter-propagating front waves meet, they will pass through each other, and upon separation, each pulse retains its original shape and amplitude. Over the course of an interaction, both constructive and destructive interference (Figs 1 and 3) is evident. Front-wave interactions, however, are not

described by simple linear superposition. Like classic solitons, the pulses are delayed (or advanced) by their mutual interaction. This causes spatial phase shifts, made apparent by 'kinked' tracks, as highlighted in Fig. 3c.

The continuum theory of dynamic brittle fracture¹ is based on equating the energy flux into a near-singular region at the tip of a crack with the energy flux needed to create new surface area. In homogeneous materials the resulting equation of motion, predicting an inertia-free crack, is highly successful¹⁷. Front waves, owing to their long lifetime, provide both inertia (as a crack front now has a record of its 'history') and a mode of dissipation that is not possible within the framework of two-dimensional theories of fracture. A full three-dimensional theory of fracture must be able to incorporate these effects. Recent work^{2-4,13-15} has also suggested that the cumulative effect of numerous asperities would be to cause a crack front to continually roughen. We point out that despite this possibility of increasing roughness, the propagating nature of front waves ensures that the cumulative perturbation history is, over time, evenly distributed throughout the front. Thus, front waves serve as an important mechanism to ensure coherence of cracks in these materials. $\square$

Methods

The experiments were conducted using soda-lime glass sheets of size $380 \times 440 \times 3\text{ mm}$ in the x (propagation), y (loading) and z (sample thickness) directions, respectively. All samples were loaded quasistatically as in ref. 17. The crack velocity, with a resolution of order 50 m s^{-1} , was measured at the sample surfaces ($z = 0$ and $z = 3\text{ mm}$ planes) as in ref. 17. The value of V_R , obtained by direct measurement, is $3,300\text{ m s}^{-1}$. Surface amplitude measurements were performed using a modified Taylor-Hobson (Surtronic 3+) scanning profilometer with spatial resolution of $0.5\text{ }\mu\text{m}$ and a $0.01\text{ }\mu\text{m}$ resolution in height measurement. Asperities were introduced in the crack path either by scribing thin lines (locally decreasing fracture energy) or filling scribed lines with superglue adhesive (locally increasing fracture energy) at the sample edges ($z = 0$ or $z = 3\text{ mm}$ planes) in the y direction. Above $v = 0.42V_R$, microbranches would spontaneously nucleate, effectively increasing the local value of the fracture energy, and thereby generating front waves. Pulse profiles (see Fig. 4) were constructed using profilometer measurements along the propagation direction in regions where the velocity was constant. These correspond to profiles formed in time at a constant position along the front and match profiles obtained in the transverse (z) direction corresponding to profiles taken at constant time.

Received 18 October 2000; accepted 8 January 2001.

- Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge, New York, 1990).
- Ben Zion, Y. & Morrissey, J. A simple re-derivation of logarithmic disordering of a dynamic planar crack due to small random heterogeneities. *J. Mech. Phys. Solids* **43**, 1363–1368 (1995).
- Perrin, G. & Rice, J. R. Disorder of a dynamic planar crack front in a model elastic medium of randomly variable toughness. *J. Mech. Phys. Solids* **42**, 1047–1064 (1994).
- Rice, J. R., Ben Zion, Y. & Kim, K. S. 3-Dimensional perturbation solution for a dynamic planar crack moving unsteadily in a model elastic solid. *J. Mech. Phys. Solids* **42**, 813–843 (1994).
- Mandelbrot, B. B. & Pasaio, D. E. Fractal character of fracture surfaces of metals. *Nature* **308**, 721–722 (1984).
- Maloy, K. J., Hansen, A., Hinrichsen, E. L. & Roux, S. Experimental measurements of the roughness of brittle cracks. *Phys. Rev. Lett.* **68**, 213–215 (1992).
- Daguier, P., Henaux, S., Bouchaud, E. & Creuzet, F. Quantitative analysis of a fracture surface by atomic force microscopy. *Phys. Rev. E* **53**, 5637–5642 (1996).
- Daguier, P., Nghiem, B., Bouchaud, E. & Creuzet, F. Pinning and depinning of crack fronts in heterogeneous materials. *Phys. Rev. Lett.* **78**, 1062–1065 (1996).
- Bouchaud, J. P., Bouchaud, E., Lapasset, G. & Planes, J. Models of fractal cracks. *Phys. Rev. Lett.* **71**, 2240–2243 (1993).
- Kulander, B. R., Barton, C. C. S. & Dean, L. *The Application of Fractography to Core and Outcrop Fracture Investigations* (The US Department of Energy, Morgantown, 1979).
- Bahat, D. *Tectono-fractography* (Springer, Berlin, 1991).
- Sharon, E. & Fineberg, J. Microbranching instability and the dynamic fracture of brittle materials. *Phys. Rev. B Cond. Matter* **54**, 7128–7139 (1996).
- Ramanathan, S. & Fisher, D. S. Dynamics and instabilities of planar tensile cracks in heterogeneous media. *Phys. Rev. Lett.* **79**, 877–880 (1997).
- Morrissey, J. W. & Rice, J. R. Crack front waves. *J. Mech. Phys. Solids* **46**, 467–487 (1998).
- Morrissey, J. W. & Rice, J. R. Perturbative simulations of crack front waves. *J. Mech. Phys. Solids* **48**, 1229–1251 (2000).
- Willis, J. R. & Movchan, A. B. Dynamic weight-functions for a moving crack. 1. mode-I loading. *J. Mech. Phys. Solids* **43**, 319–341 (1995).
- Sharon, E. & Fineberg, J. Confirming the continuum theory of dynamic brittle fracture for fast cracks. *Nature* **397**, 333–335 (1999).
- Wallner, H. Linienstrukturen an Bruchfächen. *Z. Physik* **114**, 368–378 (1939).
- Hull, D. & Beardmore, P. Velocity of propagation of cleavage cracks in tungsten. *Int. J. Fract.* **2**, 468–487 (1966).
- Willis, J. R. & Movchan, A. B. Three-dimensional dynamic perturbation of a propagating crack. *J. Mech. Phys. Solids* **45**, 591–610 (1997).

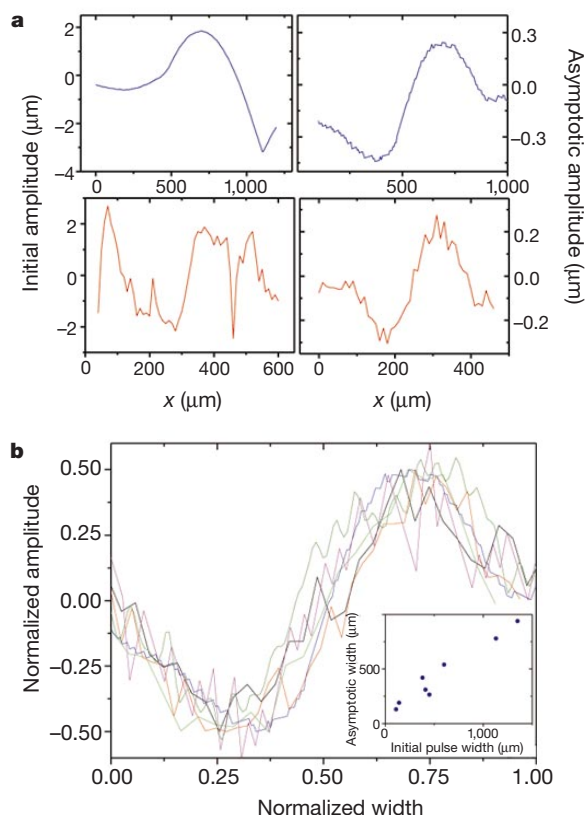


Figure 4 Asymptotically, front waves attain a characteristic, highly localized (soliton-like) profile. **a**, Comparison of initial and asymptotic profiles of front waves initiated by (top) an external asperity and (bottom) nucleation of a microbranch¹². Although the initial conditions and scales vary considerably, front waves rapidly converge to an identical form. **b**, Front waves are nonlinear entities with a well defined, scale-independent, spatial profile. Six normalized front waves used were both in steady state propagation as well as in their decay phase. The scale of the front waves is set (inset) by the width of the initial perturbation.

21. Movchan, A. B., Gao, H. & Willis, J. R. On perturbations of plane cracks. *Int. J. Solids Struct.* **35**, 3419–3453 (1998).
 22. Remoissenet, M. *Waves Called Solitons* (Springer, Berlin, 1996).

Acknowledgements

The authors acknowledge the support of the United States–Israel Binational Fund.

Correspondence and requests for materials should be addressed to J.F.
 (e-mail: jay@vms.huji.ac.il).

Orbit-related long-term climate cycles revealed in a 12-Myr continental record from Lake Baikal

Kenji Kashiwaya*, Shinya Ochiai†, Hideo Sakai† & Takayoshi Kawai‡

* Department of Earth Sciences, Kanazawa University, Kakuma, Kanazawa 920-1192, Japan

† Department of Earth Sciences, Toyama University, Toyama 930-8555, Japan

‡ National Institute for Environmental Studies, Tsukuba, Ibaragi 350-0044, Japan

Quaternary records of climate change from terrestrial sources, such as lake sediments^{1,2} and aeolian sediments^{3,4}, in general agree well with marine records^{5,6}. But continuous records that cover more than the past one million years were essentially unavailable until recently, when the high-sedimentation-rate site of Lake Baikal was exploited^{1,2,7,8}. Because of its location in the middle latitudes, Lake Baikal is highly sensitive to insolation changes⁹ and the entire lake remained uncovered by ice sheets throughout the Pleistocene epoch, making it a valuable archive for past climate. Here we examine long sediment cores from Lake Baikal that cover the past 12 million years. Our record reveals a gradual cooling of the Asian continental interior, with some fluctuations. Spectral analyses reveal periods of about 400 kyr, 600 kyr and 1,000 kyr, which may correspond to Milankovitch periods (reflecting orbital cycles). Our results indicate that changes in insolation were closely related to long-term environmental variations in the deep continental interior, over the past 12 million years.

Since Hays *et al.* found¹⁰ variations with periods of 100 kyr, 40 kyr and 20 kyr in cores taken from ocean sediments, there have been many reports of long-term global climate changes with these periods. Such variations are thought to be responses to changes in insolation: the periods of about 100 kyr, 40 kyr and 20 kyr are attributed respectively to changes in the eccentricity of the Earth's orbit, the obliquity of tilt of the Earth's axis, and the precession of its axis of rotation. Longer periods (at about 400 kyr, 600 kyr, and so on) in the insolation are known from calculated results, but they have not been much discussed. One reason for this is the lack of long, detailed climate data series. Another is that these longer-period variations are relatively small, and have been little considered when accounting for long-term global changes. Recently, the 400-kyr period has been reported to occur in the sediment record from the ocean¹¹, and also in Lake Baikal sediments^{12,13}.

Sediments from Lake Baikal have been used to study palaeoclimatic⁸ as well as modern¹⁴ environmental changes. Limited analyses of the physical properties of core BDP96—from Academician ridge in Lake Baikal, collected in 1996^{12,13}—show climatolimnological signals preserved in sediments over the past 5.0 Myr. The 400-kyr period of eccentricity and other orbital cycles are clearly exhibited in BDP96 data sets, including the 100-kyr period of eccentricity, the 40-kyr period of obliquity, and the 20-kyr precessional period, which are better known in oceanographic¹⁰

and lacustrine¹⁵ records. In addition, major shifts in oscillations (cooling) occurred around 1.0, 1.8 and 2.7 Myr ago, over periods of about 1.0 Myr, which may have been related to solar insolation and/or changes in palaeomagnetism^{13,16}.

In spring 1998, we obtained long cores (BDP98, 600 m) from Academician ridge (53° 44' 40" N, 108° 24' 30" E in central Lake Baikal), at nearly the same site as core BDP96. Academician ridge is thought to be a convenient site for obtaining a continuous record because it is a topographically isolated ridge, little influenced by fluvial input or turbidity flows. Core BDP98 spans approximately the past 12.0 Myr, based on palaeomagnetic correlations (H.S. and S. Nomura, unpublished observations; see Supplementary Information). The tentative age scale used here is based on geomagnetic polarity reversals (Fig. 1a; ages from ref. 17). Between the reversals, linear interpolation is used for age scaling. We excluded from statistical analyses the core segment between subchrons C3An.2n and C3Bn (6.567–6.935 Myr ago), owing to sample distortion (seemingly related to faults) that would have erroneously suggested abnormally high sedimentation rates. Analytical results from the upper 200 m of the cores are nearly identical to those from core BDP96. The present study is mainly concerned with physical

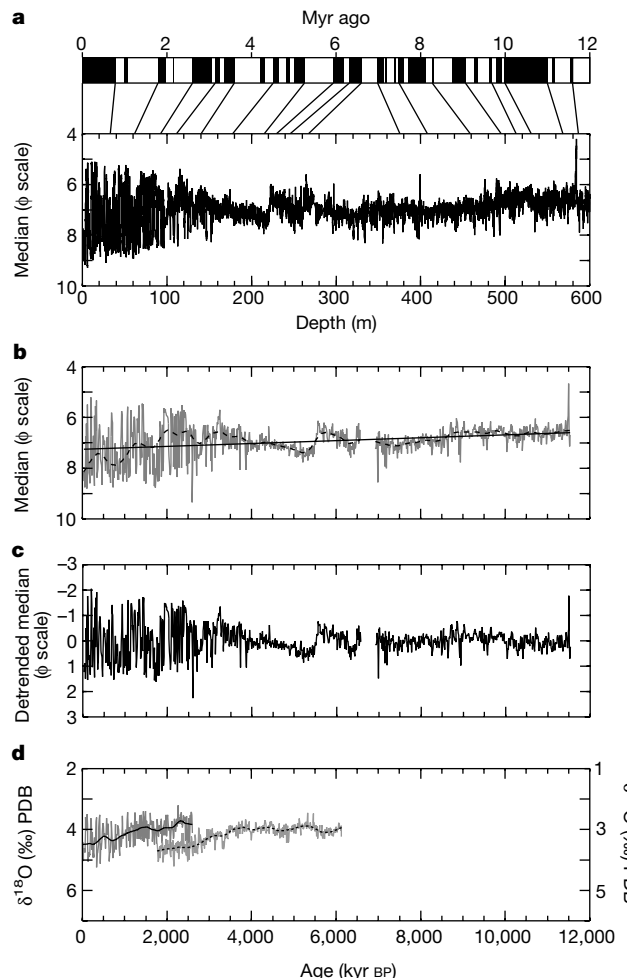


Figure 1 Long-term fluctuations discussed in the text. **a**, Reference geomagnetic polarity timescale¹⁷ (top), and our data showing mean grain size (bottom); solid lines mark correlations between magnetostratigraphy and the core. **b**, Equally spaced mean grain size (light solid curve) based on the tentative age scale described in the text (dotted line, weighted smoothing³⁰; heavy solid line, regression). **c**, Detrended mean grain size. **d**, Oceanic $\delta^{18}\text{O}$ records from ODP sites. Site 677, left, with solid curve (weighted smoothing); Site 846, right, with dotted curve (weighted smoothing)^{6,28}.

properties, especially mean grain size, and subsamples for this purpose were taken every 5 cm along the core, amounting to some 11,000 subsamples. Mean grain sizes were measured with a laser reflection analyser, and are shown in Fig. 1b. Diatoms are the principal component of biogenic productivity in the sediments from Academician ridge¹, and closely reflect changes in physical properties such as grain size for at least the past 5.0 Myr (ref. 13). Grain size was comparatively large during warm periods and small during cold periods, and has accordingly been used as a proxy for climato-limnological fluctuations^{12,13}.

We obtained equally spaced data points (10-kyr intervals) for statistical analysis by interpolation for the intervals 0–6.5 and 7.0–11.5 Myr ago. Spectral analysis (Maximum Entropy Method)¹⁸ was used for both time intervals to calculate periods longer than 350 kyr. As shown in Fig. 2a and b, there were periods of about 1,000 kyr, 600 kyr and 400 kyr in both time intervals, even if their amplitudes were low between 7.0 and 11.5 Myr ago. The 1,000- and 400-kyr periods are present in the shorter core obtained from nearly the same site, indicating that the 400-kyr period is related to the eccentricity parameter of solar insolation, while the 1,000-kyr period needs more data to verify its existence¹³. We found another long period at about 600 kyr (although a little broad and fluctuating), which may also be related to the eccentricity parameter (604-kyr period)¹⁹, even though the variations in eccentricity are small. Cross-spectral analysis²⁰ of mean grain size and insolation (at 65° N in July, calculated in ref. 21) was introduced for the two time intervals; the results shown in Fig. 2e and f suggest that there is some relationship between the grain size and insolation at those periods, especially at about 1,000 kyr and 400 kyr.

Filter analysis has been used²² to check temporal changes in periodicity (Fig. 3); we detrended our data before applying this

technique (Fig. 1c). The 150- to 80-kyr band-pass filter addresses periods around 100 kyr, which are also related to another eccentricity parameter. A period of about 1,000 kyr is clearly seen in both the intervals studied (0–6.5 and 7.0–11.5 Myr ago), even though there is some distortion at about 4 Myr ago, and the amplitude was much smaller before 6.5 Myr ago (Fig. 3a). The periodicity evidently continued during the whole period, despite a lack of data between about 6.5 and 7.0 Myr ago. A period of about 600 kyr is clearly seen in the whole interval studied, although there were shifts in amplitude at about 7.0, 3.0 and 1.0 Myr ago (Fig. 3b), and a gradual increase, especially between 2.8 and 1.2 Myr ago. An approximate 400-kyr period related to the eccentricity parameter of insolation is also seen in the whole interval studied (Fig. 3c). The amplitude was small before 4.0 Myr ago and increased afterwards, especially at 3.5–1.5 Myr ago, as also seen in the shorter core¹³. Periods around 100 kyr are clearly seen after 3.0–2.7 Myr ago and show a stable cycle after 1.0 Myr ago. Shifts at about 7.0–6.5, 4.0, 3.0–2.7 and 1.0 Myr ago in the amplitude of the periods seem to correspond to global climate shifts, as discussed below.

We now examine temporal changes in periodicity in more detail; to do this, we consider five intervals of time, and use spectral analysis¹⁸ to find periods—and harmonic analysis to check periods—within these intervals (Fig. 4). In the interval of 0–3.0 Myr ago, periods at about 1,000, 600 and 400 kyr have large amplitudes. Periods similar to these are also present at 2.0–5.0 Myr ago and 4.0–6.5 Myr ago, although their values fluctuate slightly and their amplitudes gradually decline. They become smaller in the intervals of 7.0–10.0 and 9.0–11.5 Myr ago, although the ~1,000-kyr period is comparatively large. These also support the suggestion that periods at about 1,000, 600 and 400 kyr may exist in all the intervals.

Periods of around 400 and 100 kyr have been reported, along with the 1,000-kyr period, in Lake Baikal sediments^{12,13}. The 400- and 100-kyr periods are related to eccentricity parameters discussed previously^{12,13}, but the 1,000-kyr and 600-kyr periods found here have not been clearly discussed. Results of spectral analysis of the oceanic records (from ODP sites 677 and 864) also suggest that similar periods are found in both of them (Fig. 2c and d). The intervals of the records roughly correspond to the two above-

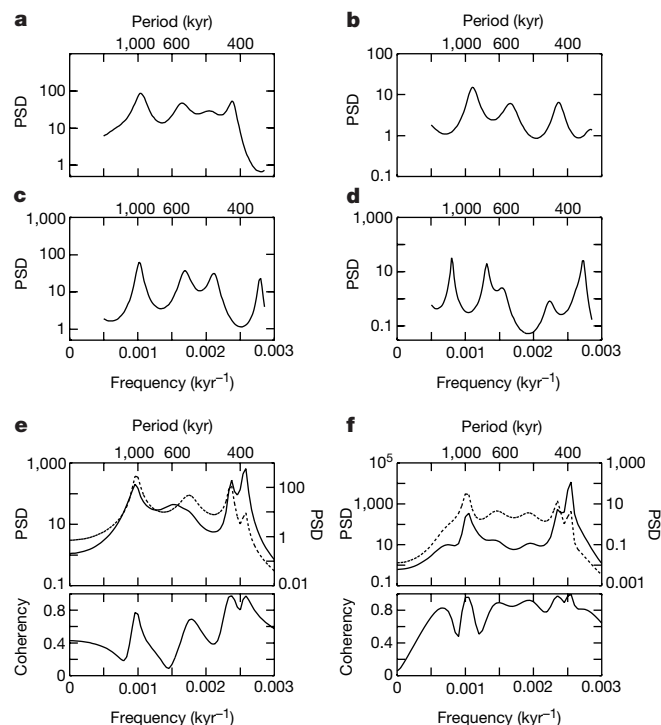


Figure 2 Dominant periods for the long-term fluctuations. **a–d**, Spectral density from spectral analysis; 0–6.5 Myr ago (**a**), 7.0–11.5 Myr ago (**b**), for mean grain size of BDP98; **c**, ODP results ($\delta^{18}\text{O}$ data from site 677) and **d**, ODP results ($\delta^{18}\text{O}$ data from site 864). **e, f**, Cross-spectral comparison of insolation²¹ (solid lines, left axis) and mean grain size (dotted lines, right axis); **e**, 0–6.5 Myr ago, and **f**, 7.0–11.5 Myr ago. A band-pass filter of period 2,000–350 kyr is used. PSD in the ordinate (log unit) indicates power spectral density.

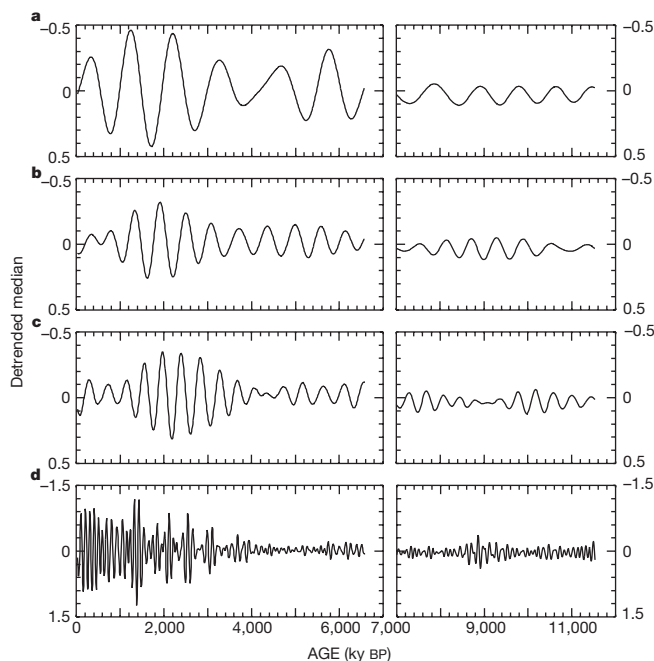


Figure 3 Filtered curves for mean grain size of BDP98. Band-pass filters of the following periods were used: **a**, 1,500–750 kyr; **b**, 750–500 kyr; **c**, 500–350 kyr; and **d**, 150–75 kyr.

mentioned ones (0–3.0 Myr and 2.0–5.0 Myr ago, Fig. 4a and b, respectively), although the 400-yr period is split in both cases and the 1,000-kyr period is a little longer. A period of 700 kyr appears in addition to the 600-kyr period in the record from Site 864. The periods are related to the eccentricity, because spectral analysis of the eccentricity data calculated in ref. 21 over the past 12 Myr indicates that there exist periods at about 1,000, 600–700 and 400–500 kyr, which is supported by the data in ref. 23. The 1,000-kyr period may represent a beat of periods in the climatic precession (23,708 and 23,149 yr)²⁴, or be a nonlinear combination of different orbital parameters (for example, a beat of periods in the obliquity; a 1,190-kyr period could be generated by the 41,090-yr and 39,719-yr periods reported in ref. 24). The 600–700-kyr period may be related to the above-mentioned period in the eccentricity. It may also be a beat of climatic precession periods²⁵. These considerations suggest that all the periods found in the present work are related to periods contained in the insolation.

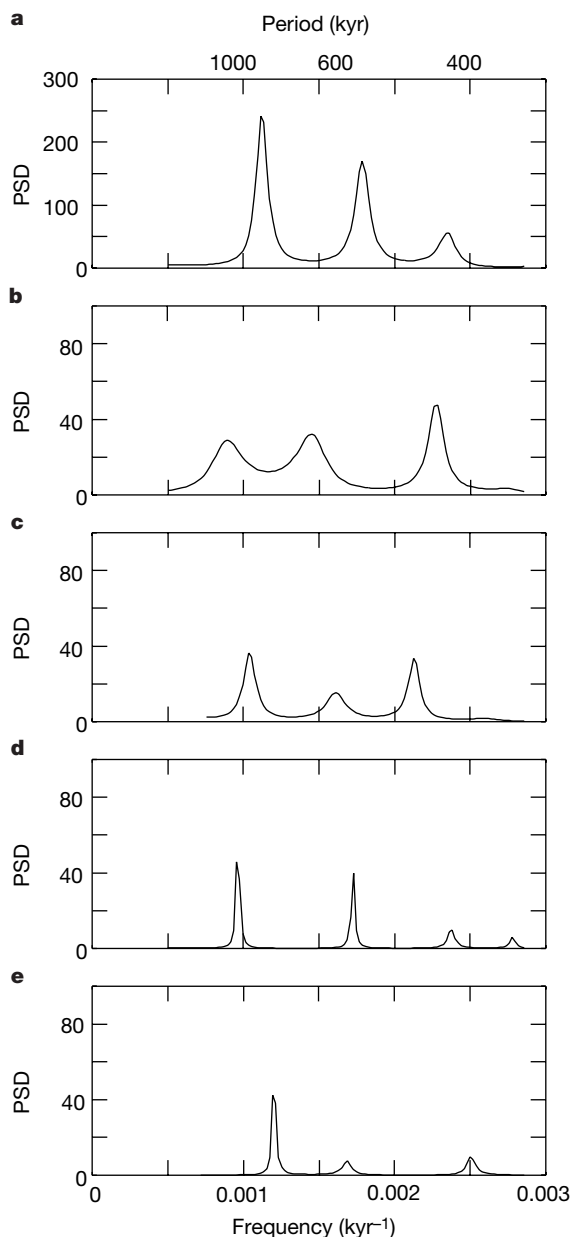


Figure 4 Results of spectral analysis with a band-pass filter of period 2,000–350 kyr for five intervals. **a**, 0–3 Myr ago; **b**, 2–5 Myr ago; **c**, 4–6.5 Myr ago; **d**, 7–10 Myr ago; and **e**, 9–11.5 Myr ago. PSD in the ordinate indicates power spectral density.

Our data suggest gradual continental cooling and much larger fluctuations after ~3.0 Myr ago, coinciding with global climate change over the past 12 Myr (Fig. 1b). A shift at about 8.5–8.8 Myr ago may be related to the uplift of the Himalayas and the Tibetan plateau, which initiated the Indian monsoon^{26,27}; the 1,000-kyr period strengthened after this shift. A ‘rewarming’ period at about 5.5–6.0 Myr ago coincides with a ‘diminution of upwelling’ in the Indian Ocean²⁷. The amplitude of fluctuations began to increase at about 4.0 Myr ago, especially in the 400-kyr and 100-kyr periods, suggesting another environmental shift at this time. The even higher amplitudes found in the interval 2.6–2.8 Myr ago may reflect the expansion of ice sheets at about 2.7 Myr ago^{3,5}; after this interval, the 600-kyr and 100-kyr periods become more distinct. The interval 1.7–1.9 Myr ago may relate to the environmental change once defined as the beginning of the Quaternary^{6,28}. The interval 0.8–1.1 Myr ago may signal the initiation of full glaciation in the mid-Pleistocene²⁹; the 100-kyr period has become more cyclic since this interval. These shifts suggest that the establishment of the ‘monsoon regime’ and the intensification of glaciation in the Northern Hemisphere may have made signals of the orbit-related periods more sensitive to global climate change. □

Received 15 May; accepted 19 December 2000.

- Colman, S. M. *et al.* Continental climate response to orbital forcing from biogenic silica records in Lake Baikal. *Nature* **378**, 769–771 (1995).
- Kashiwaya, K. *et al.* Orbital signals found in physical and chemical properties of bottom sediments from Lake Baikal. *J. Paleolimnol.* **14**, 293–297 (1997).
- Kukla, G. Loess stratigraphy in central China. *Quat. Sci. Rev.* **6**, 191–219 (1987).
- Maher, B. & Thompson, R. Mineral magnetic record of the Chinese loess and paleosols. *Geology* **19**, 3–6 (1991).
- Shackleton, N. J. *et al.* Oxygen isotope calibration of the onset of ice rafting and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
- Shackleton, N. J., Hall, M. A. & Pate, D. Pliocene stable isotope stratigraphy of site 846. *Proc. ODP Sci. Res.* **138**, 337–355 (1995).
- Baikal Drilling Project Members (BDP93). Preliminary results of the first scientific drilling on Lake Baikal, Buguldeika site, southeastern Siberia. *Quat. Int.* **37**, 3–17 (1997).
- Baikal Drilling Project Leg II Members (BDP96). Continuous continental paleoclimate record for the last 4.5 to 5 million years revealed by leg II of Lake Baikal scientific drilling. *Eos* **78**, 597 (1997).
- Short, D. A., Mengel, J. G., Crowley, T. J., Hyde, W. T. & North, G. R. Filtering of Milankovitch cycles by Earth's geography. *Quat. Res.* **35**, 157–173 (1991).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Clemens, S. C. & Tiedemann, R. Eccentricity forcing of Pliocene–Early Pleistocene climate revealed in a marine oxygen-isotope record. *Nature* **385**, 801–804 (1997).
- Kashiwaya, K., Ryugo, M., Sakai, H. & Kawai, T. Long-term climato-limnological oscillation during the past 2.5 million years printed in Lake Baikal sediments. *Geophys. Res. Lett.* **25**, 659–663 (1998).
- Kashiwaya, K., Sakai, H., Ryugo, M., Horii, M. & Kawai, T. Long-term climato-limnological cycles found in a 3.5-million-year continental record. *J. Paleolimnol.* (in the press).
- Flower, R. J. Paleolimnology and recent environmental change in Lake Baikal: an introduction and overview of interrelated concurrent studies. *J. Paleolimnol.* **20**, 107–117 (1998).
- Kashiwaya, K., Fukuyama, K. & Yamamoto, A. Time variations in coarse materials from lake bottom sediments and secular paleoclimatic change. *Geophys. Res. Lett.* **18**, 1245–1248 (1991).
- Kashiwaya, K., Tanaka, A., Sakai, H. & Kawai, T. Paleoclimatic signals printed in Lake Baikal sediments. In *Lake Baikal* (eds Minoura, K. *et al.*) (Amsterdam, Elsevier, 2001).
- Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).
- Barrodale, I. & Ericksson, R. E. Algorithm for least-square linear prediction and maximum entropy spectral analysis, part 1. Theory. *Geophysics* **45**, 420–432 (1980).
- Berger, A. Long-term variations of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Hayashi, Y. Space-time cross spectral analysis using the maximum entropy method. *J. Meteorol. Soc. Jpn* **59**, 620–624 (1981).
- Laskar, J., Joutel, F. & Boudin, F. Orbital, precessional, and insolation quantities for the Earth from –20 Myr to +10 Myr. *Astron. Astrophys.* **270**, 522–533 (1993).
- Ormsby, J. F. A. Design of numerical filters with applications to missile data processing. *J. Assoc. Comput. Mech.* **8**, 440–466 (1961).
- Lourens, L. J. & Hilgen, F. J. Long-periodic variations in the Earth's obliquity and their relation to third-order eustatic cycles and late Neogene glaciations. *Quat. Int.* **40**, 43–52 (1997).
- Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
- Berger, A. & Loutre, M. F. Origine des fréquences des éléments astronomiques intervenant dans le calcul de l'insolation. *Bulletin de la Classe des Sciences, Académie Royale Belgique, 6^e série, Tome 1*, 45–106 (1990).
- Prell, W. L. & Kutzbach, J. E. Sensitivity of the Indian monsoon to forcing parameters and implication for its evolution. *Nature* **360**, 647–652 (1992).
- Kroon, D., Steens, T. & Troelstra, S. R. Onset of monsoonal related upwelling in the western Arabian Sea as revealed by planktonic foraminifers. *Proc. ODP Sci. Res.* **117**, 257–263 (1991).

28. Shackleton, N. J., Berger, A. & Peltier, W. R. An alternative astronomical calibration of the lower Pleistocene timescale based on ODP site 677. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251–261 (1990).
29. Maasch, K. A. Statistical detection of the mid-Pleistocene transition. *Clim. Dyn.* **2**, 133–143 (1988).
30. Chambers, J., Cleveland, W., Kleiner, B. & Tukey, P. *Graphical Methods for Data Analysis* 395 (Chapman & Hall/CRC, Boca Raton, 1983).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the BDP Leg IV members from Russia, America and Japan who worked to drill the BDP98 sediment cores, and J. Laskar, A. Berger and M.-F. Loutre for providing the calculated results on insolation. We also thank colleagues in the hydro-geomorphological laboratory at Kanazawa University and the geomagnetic laboratory of Toyama University for their support in this research.

Correspondence and requests for materials should be addressed to K.K. (e-mail: kashi@kenroku.kanazawa-u.ac.jp).

Earthquake slip on oceanic transform faults

Rachel E. Abercrombie & Göran Ekström

Department of Earth and Planetary Sciences, Harvard University,
20 Oxford Street, Cambridge, Massachusetts 02138, USA

Oceanic transform faults are one of the main types of plate boundary, but the manner in which they slip remains poorly understood. Early studies suggested that relatively slow earthquake rupture might be common^{1,2}; moreover, it has been reported that very slow slip precedes some oceanic transform earthquakes, including the 1994 Romanche earthquake^{3–5}. The presence of such detectable precursors would have obvious implications for earthquake prediction. Here we model broadband seismograms of body waves to obtain well-resolved depths and rupture mechanisms for 14 earthquakes on the Romanche and Chain transform faults in the equatorial Atlantic Ocean. We found that earthquakes on the longer Romanche transform are systematically deeper than those on the neighbouring Chain transform.

These depths indicate that the maximum depth of brittle failure is at a temperature of $\sim 600^\circ\text{C}$ in oceanic lithosphere. We find that the body waves from the Romanche 1994 earthquake can be well modelled with relatively deep slip on a single fault, and we use the mechanism and depth of this earthquake to recalculate its source spectrum. The previously reported slow precursor can be explained as an artefact of uncertainties in the assumed model parameters.

The depth extent of seismic slip in shallow oceanic earthquakes is poorly known, limiting our understanding of how oceanic transform faults move, and how the lithospheric composition affects the earthquake rupture process. Estimates of seismic coupling determined from the ratio of seismic slip to plate motion by summing the earthquake moments are crucially dependent on the width of the fault assumed to slip in earthquakes. Oceanic crust is only $\sim 6\text{ km}$ thick. Theoretical strength profiles predict that seismic slip should extend into the upper mantle, unlike continents, where it is restricted to the upper crust⁶.

The international agencies locating earthquakes typically report the depth of earthquakes on oceanic ridges and transforms as 10 or 15 km. More accurate depths and mechanisms of oceanic earthquakes have been obtained by modelling long period (LP) body waves^{7–9}. The depths of oceanic intraplate earthquakes increase with lithospheric age, implying that temperature controls the depth of the brittle–ductile transition⁷. The LP data were inadequate to resolve similar trends in the depths of earthquakes along the relatively young transform faults. Few transform fault studies have used ocean-bottom seismometers, and most were unable to resolve hypocentres accurately¹⁰.

Very slow slip occurs in some earthquakes¹¹, but the observations of oceanic transform earthquakes^{3–5} remain controversial. This is because the hypothesized slow component is often a small fraction of the total moment and because other authors have failed to confirm the proposed slow rupture during the largest earthquake considered^{12,13}. The shape of the long-period spectrum, used to identify possible slow rupture, is strongly dependent on the Earth structure and source parameters (including depth) assumed in the modelling¹⁴. More accurate measurements of these parameters are therefore necessary.

Modern broadband data have significantly higher resolution than the LP data. We use these data to compare earthquakes on two contrasting transform faults on the equatorial Mid-Atlantic Ridge.

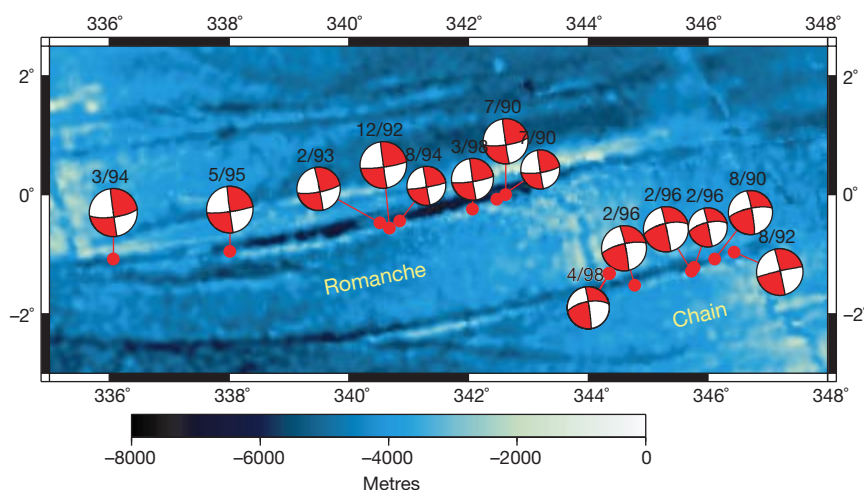


Figure 1 Map of the equatorial Mid-Atlantic Ridge, showing the Romanche and Chain transform faults and also the bathymetry^{26,27}. The transform valleys are dark blue and the interconnecting ridges are pale blue. The focal mechanisms obtained in the broadband body-wave modelling are joined to their National Earthquake Information Center hypocentre locations (red circles), with radius proportional to magnitude. The earthquakes

on the Romanche transform all dip to the south at $\sim 80^\circ$, and those on the Chain transform all dip to the north at $\sim 73^\circ$. The centroid-moment tensor (CMT) mechanisms are not accurate enough to resolve these differences. The active faults thus seem planar, but the reason for their opposite sense of dip is unknown.

We compare earthquakes on the Romanche transform, which is one of the longest on the mid-ocean ridge system and offsets 50-Myr lithosphere, with events on the neighbouring Chain transform, which is shorter and offsets younger lithosphere (17 Myr) (Fig. 1). We use broadband body waves to determine focal mechanisms and depths for the 14 largest earthquakes in the Harvard CMT catalogue

since 1990 (refs 15, 16) (Figs 1 and 2). Most of these earthquakes are well recorded, and depths and mechanisms are well constrained.

The mechanisms of the earthquakes on each transform are extremely similar. They all have strikes within 2° of the strike of the transforms ($\sim 80^\circ$), implying that the transform faults are planar. Using these well-constrained focal mechanisms, we determine the distribution of slip in the largest, 1994 Romanche earthquake (Fig. 3). This earthquake has previously been reported to have a complex rupture, including a slow precursor⁴. Strong eastward directivity is clear in the waveforms. The earthquake had a small amplitude onset lasting about 12 s (subevent A (ref. 4), which is not unusual for a large earthquake¹⁷). The main slip occurred in two large subevents to the east of the hypocentre. Their centroid depth agrees well with that obtained from the body-wave modelling (12 km; all depths quoted are below the sea floor) and seismic slip extends to

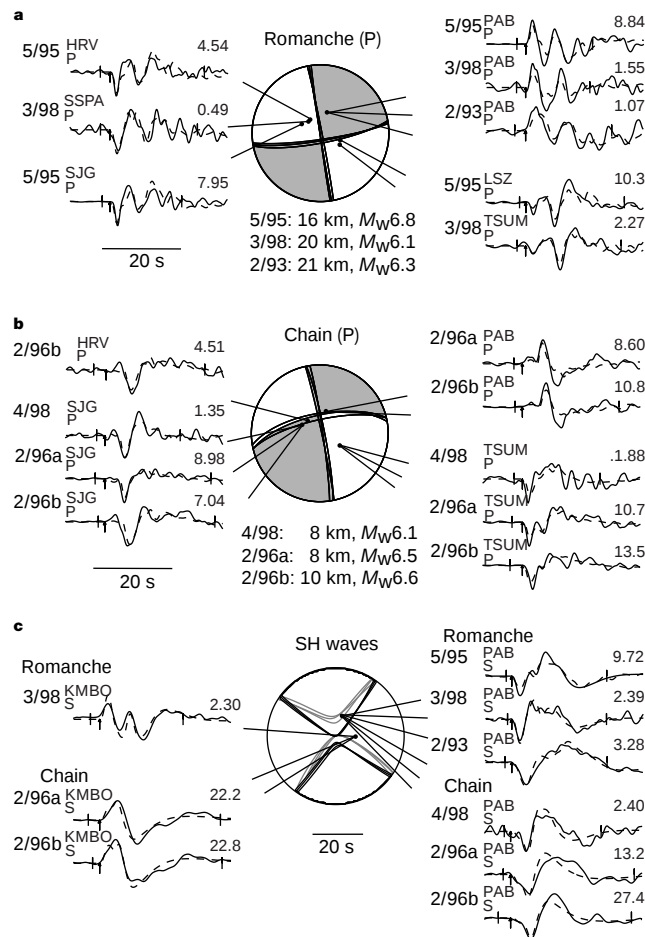


Figure 2 Comparison of the waveforms from six earthquakes, three on each of the Romanche and Chain transforms. The observed displacement broadband waveforms (solid lines, 20 samples per second), are corrected for the instrument response and filtered between 1 and 150 s. The dashed lines are the synthetics. The stations are chosen to allow comparison between earthquakes on each transform and between transforms. The dates, depths and magnitudes are given. The amplitudes of the seismograms (μm) are given at the top right. The arrows mark the P picks; the vertical lines indicate the duration used in the inversion. **a**, **b**, P waves are shown from earthquakes on the Romanche transform (**a**) and the Chain transform (**b**). All six mechanisms are shown, and those of the 1995 and February 1996b earthquakes are shaded. **c**, SH waves from all six earthquakes. The SH mechanisms of the Romanche earthquakes are shown in black, and those on Chain in grey. The seismograms from earthquakes on individual transforms are very similar but differ between transforms, constraining the different dip directions. The depths of the earthquakes are well constrained by the arrival times of the depth phases. The waveforms of the Chain earthquakes are shorter in duration than the Romanche events, implying that they are shallower. P wave arrival times are picked on velocity seismograms. The P and SH waveforms (between 9 and 36 per earthquake) are inverted for mechanism, centroid depth and moment rate²⁸. The model source structure consists of 5 km of water above a 6-km crust²⁸. Non-double couple components were initially included but were negligible, and the mechanisms are fixed to double couples. Rupture velocity is included when it significantly improves the fit. The data for two earthquakes (July 1990c, Fig. 1, and February 1996b) are insufficient to constrain the mechanisms and so they are set equal to those of neighbouring events with similar waveforms.

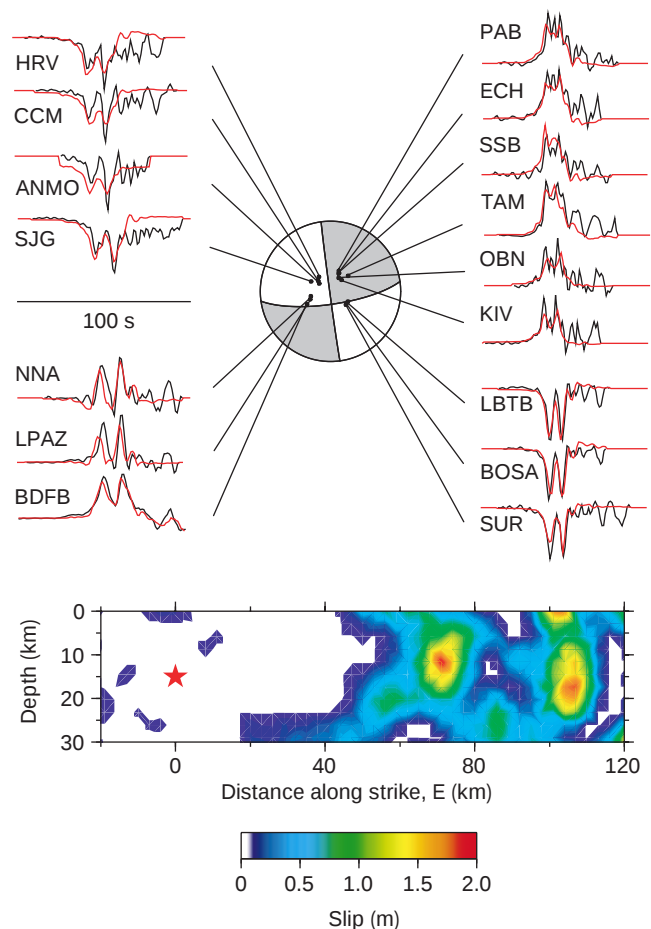


Figure 3 Inversion for slip distribution in the 14 March 1994 Romanche earthquake (M_w 7.1). We use the mechanism obtained in the body-wave modelling (strike 82° , dip 75° , rake 179°) to invert for slip on a finite fault plane. We use broadband displacement P waves, low-pass-filtered at 1 Hz. We use an oceanic structure and invert for slip on a $150 \text{ km} \times 39 \text{ km}$ fault plane divided into $3 \text{ km} \times 3 \text{ km}$ subfaults, extending from the sea floor (0 km)²⁹. We use fixed rupture velocities between 3.0 and 4.5 km s^{-1} , but only those between 3.5 and 4 km s^{-1} can fit the waveforms well. The observed waveforms are shown in black (labelled with station code) and the synthetics in red. The strong eastward directivity is clear when the arrival times of the first large pulse and the duration of the waveforms in the east are compared with those in the west. All the synthetic seismograms start 20 s before the predicted arrival time. The preferred model is for a rupture velocity of 4 km s^{-1} , and has a moment of $5.1 \times 10^{19} \text{ N m}$. Two large subevents are well resolved, and some small slip occurred near the hypocentre. The depth range is from near the surface to $\sim 20 \text{ km}$. Slip in the upper 5 km, and below 20 km, is not required; constraining the slip to between these depths results in an insignificant decrease in fit. The variance decrease of the preferred model is 80%.

~20 km. The first of these subevents corresponds to subevent B (ref. 4). McGuire *et al.*⁴ identified subevents by picking arrivals in the observed seismograms; they did no waveform modelling. This is difficult for strike-slip earthquakes in which many arrivals are small or nodal. From the timing of their picks, McGuire *et al.* located subevent B 80 km northeast of A, on a separate fault. This jump is over an order of magnitude greater than observed in surface ruptures¹⁸ or numerical simulations¹⁹. Our model does not need such a jump, and requiring it worsens the waveform fit. In our mechanism the stations to the SE are nodal, and those to the SW have small P amplitudes (Fig. 3). We infer that McGuire *et al.*⁴ misidentified the strong surface reflections (such as sP) as P. This would result in an overestimate of the time delay between the two subevents at the southern stations and would thus displace the location of subevent B to the north. Waveform modelling in the frequency range in which most energy is radiated is essential when investigating strike-slip earthquakes.

McGuire *et al.*⁴ also modelled the long-period source spectrum of the 1994 Romanche earthquake. They used synthetic reference seismograms calculated for an earthquake located in the 21-km crust in the Preliminary Reference Earth Model (PREM)²⁰ at a fixed depth of 15 km to obtain the source spectrum (Fig. 4). We recalculate the source spectrum using an oceanic crustal model at the source, and determine the mechanism and depth independently

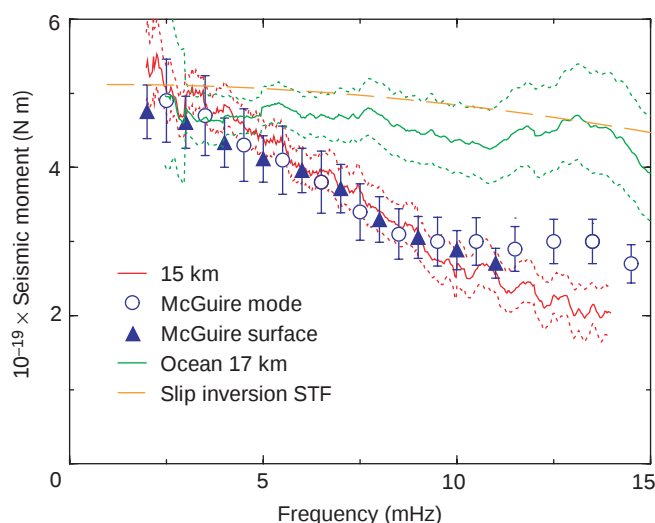


Figure 4 Long-period source spectra of the 1994 Romanche transform earthquake (M_w 7.0). The blue symbols are the spectrum obtained by McGuire *et al.*⁴ from vertical Rayleigh waves by using the PREM²⁰ structure at the source, the Harvard CMT mechanism, and a depth of 15 km below the surface of the model. Note the increase in amplitude at frequencies below ~8 mHz, where the spectrum of a typical earthquake of this magnitude should be almost flat. It is this apparent excess long-period energy that is taken as evidence for a slow rupture component^{3–5}. We obtain the source spectrum in a similar way, also using vertical Rayleigh waves. We calculate the synthetic seismograms with a hybrid method. Fundamental mode surface waves are synthesized with the use of WKBJ ray-theory, incorporating the excitation in an oceanic elastic structure. Overtones are calculated with normal-mode summation. The spectrum of the observed seismogram is divided by that of the synthetic at each station, and then all the stations are averaged to obtain the mean source spectrum. The red spectrum used synthetics calculated for 15 km depth in PREM²⁰ crustal structure and the Harvard catalogue CMT mechanism. The dotted lines are the 95% confidence limits of the mean. The results agree with those of McGuire *et al.*⁴. The green source spectrum is calculated using the mechanism and depth determined independently from the body waves (Fig. 3), and a more appropriate oceanic velocity structure at the source (5 km of water above a 6.5 km oceanic crust and PREM²⁰ mantle). The orange dashed line is the spectrum of the source time function (STF, 33 s duration) obtained from the slip inversion. The recalculated source spectrum (green) is consistent with that from the higher-frequency body waves, and shows no evidence of an increase in amplitude at long periods.

from the body-wave modelling. The resulting spectrum is flat at long periods with no resolvable slow component (Fig. 4). The recalculated source spectrum is in good agreement with the source spectrum predicted in the slip inversion of the higher-frequency body waves. This suggests that the proposed slow component is an artefact of using an inadequate source model. A small precursory ramp in the highly filtered time-domain records has also been claimed as supporting evidence for precursory slip⁴, but the amplitude is so small that these observations are ambiguous.

The depths of all 14 earthquakes obtained in the broadband body-wave modelling are shown in Fig. 5. The systematic difference between the two transforms is clear. The earthquakes on Romanche have centroid depths between 7 and 20 km, and those on Chain between 3 and 11 km. The Romanche transform offsets relatively old, cold lithosphere, and the systematic difference implies that temperature controls the maximum depth of seismic slip on oceanic transform faults as it does elsewhere^{6,7}. The centroids are located in a similar temperature range on both transforms. We can estimate the rupture area of the earthquakes from slip inversions (Figs 3 and 5) and from global relationships between earthquake moment and rupture area²¹. The 600 °C isotherm thus seems a good estimate of the limit of seismic slip. This is consistent with previous results for intraplate oceanic earthquakes⁷ but is cooler than previously estimated for transform fault events⁹. The depth of the isotherms can be

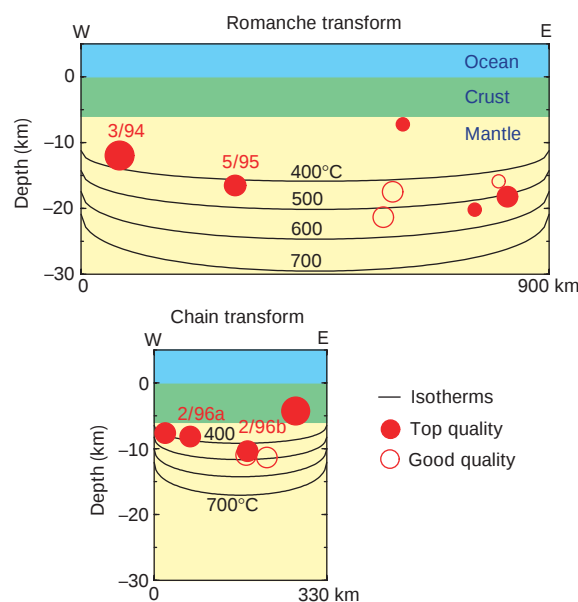


Figure 5 Depth and temperature of earthquakes on the Romanche and Chain transforms. The thickness of oceanic crust is about one-sixth that of average continental crust. The centroid depths of the earthquakes studied are plotted as red circles with symbol size proportional to magnitude. Solid symbols represent earthquakes with the best-resolved parameters, and open symbols those with fewer data or poorer fits. All the depths are accurate to within ~3 km. Isotherms, calculated by using a half-space cooling model and averaging both sides of the transform, are plotted in black³⁰. Slip inversions of the 1994 and 1995 Romanche earthquakes suggest that the former ruptured from near the surface to ~20 km depth, and the latter from ~10 to 25 km depth. The 600 °C isotherm therefore seems a reasonable estimate of the temperature controlling the maximum depth of seismic slip on these transforms. The 1992 Chain earthquake is the only event with a centroid within the crust, but it was large enough to have ruptured into the upper mantle. We sum the moments of all the predominantly strike-slip earthquakes in the 25-yr Harvard CMT catalogue along the Romanche and Chain transforms. Assuming a rigidity of 3×10^{10} N m⁻², and assuming that the transforms have lengths of 900 and 330 km and widths of 20 and 13 km, respectively, we obtain a seismic slip rate of 17 mm yr⁻¹ on Romanche and 20 mm yr⁻¹ on Chain. The seismic slip during this period might not be representative of the longer-term average, but it is higher than that calculated for the Romanche transform by Brune²⁴ for the period 1920–52.

varied by using different thermal models²². The depths of the earthquakes depend on the assumed crustal thickness and velocities in the upper lithosphere, but reasonable variation²³ would make a difference of no more than 1 km. A temperature of 600 °C is approximate but is consistent with the onset of ductile deformation in olivine-rich materials at probable strain rates⁷. The systematic difference between the two transforms is independent of these uncertainties. The depths of the earthquakes on the Romanche transform are deeper than previously determined for oceanic transform earthquakes; they are also in older lithosphere.

The degree of seismic coupling of oceanic transform faults has remained poorly constrained since it was first studied²⁴, because the width of the seismic zone was unknown. Our large seismic widths result in relatively low seismic slip rates (Fig. 5), only about half of the 37 mm yr⁻¹ tectonic slip rate²⁵.

If we are not underestimating the seismic slip, then half of the slip must be occurring aseismically. It has been suggested that aseismic and slow (possibly precursory) slip might occur in the upper mantle, whereas higher speed seismic slip occurs in a very shallow seismogenic zone⁴. The depth extent of the seismic slip in the 1994 and 1995 earthquakes (Figs 3 and 5) and the centroid depths of the other events provide clear evidence of seismic rupture extending well into the upper mantle. Our results imply that significant aseismic slip is occurring and it is likely to be at shallow depths, perhaps within the serpentinized zone.

The depth of seismic slip on oceanic transform faults is controlled by temperature and is limited by the ~600 °C isotherm. The focal mechanisms of earthquakes of the two transform faults show impressive consistency, indicating that the faults are highly planar. We also find that the two previously identified⁴ unusual aspects of the 1994 Romanche earthquake are probably artefacts of the analysis procedure used. No detectable precursors to oceanic transform earthquakes can be resolved unambiguously with present seismic data and analysis techniques. □

Received 1 March; accepted 11 December 2000.

- Kanamori, H. & Stewart, G. S. Mode of the strain release along the Gibbs fracture zone, Mid-Atlantic Ridge. *Phys. Earth Planet. Inter.* **11**, 312–332 (1976).
- Okal, E. A. & Stewart, L. M. Slow earthquakes along oceanic fracture zones: evidence for asthenospheric flow away from hotspots? *Earth Planet. Sci. Lett.* **57**, 75–87 (1992).
- Ihmlé, P. F. & Jordan, T. M. Teleseismic search for slow precursors to large earthquakes. *Science* **266**, 1547–1551 (1994).
- McGuire, J. J., Ihmlé, P. F. & Jordan, T. H. Time domain observations of a slow precursor to the 1994 Romanche transform earthquake. *Science* **274**, 82–85 (1996).
- Ihmlé, P. F., Harabaglia, P. & Jordan, T. H. Teleseismic detection of a slow precursor to the great 1989 Macquarie Ridge earthquake. *Science* **261**, 177–183 (1993).
- Sibson, R. H. Fault zone models, heat flow and the depth distribution of earthquakes in the continental crust of the United States. *Bull. Seismol. Soc. Am.* **72**, 151–163 (1982).
- Wiens, D. A. & Stein, S. Age dependence of oceanic intraplate seismicity and implications for lithospheric evolution. *J. Geophys. Res.* **88**, 6455–6468 (1983).
- Engeln, J. E., Wiens, D. A. & Stein, S. Mechanisms and depths of Atlantic transform faults. *J. Geophys. Res.* **91**, 548–577 (1986).
- Bergman, E. A. & Solomon, S. C. Transform fault earthquakes in the North Atlantic: source mechanisms and depth of faulting. *J. Geophys. Res.* **93**, 9027–9057 (1988).
- Wilcock, W. S. D., Purdy, G. M. & Solomon, S. C. Microearthquake evidence for extensions across the Kane transform fault. *J. Geophys. Res.* **95**, 15439–15462 (1990).
- Kanamori, H. & Kikuchi, M. The 1992 Nicaragua earthquake: a slow tsunami earthquake associated with subducted sediments. *Nature* **361**, 714–716 (1993).
- Kedar, S., Watada, S. & Tanimoto, T. The 1989 Macquarie Ridge earthquake: seismic moment estimation from long period free oscillations. *J. Geophys. Res.* **99**, 17893–17907 (1994).
- Velasco, A. A., Ammon, C. J. & Lay, T. Source time function complexity of the great 1989 Macquarie Ridge earthquake. *J. Geophys. Res.* **100**, 3989–4009 (1995).
- Silver, P. G. & Jordan, T. H. Total-moment spectra of fourteen large earthquakes. *J. Geophys. Res.* **88**, 3273–3293 (1983).
- Dziewonski, A. M. & Woodhouse, J. H. An experiment in systematic study of global seismicity: centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *J. Geophys. Res.* **88**, 3247–3271 (1983).
- Dziewonski, A. M., Ekström, G. & Maternovskaya, N.-N. Centroid-moment tensor solutions for October–December. *Phys. Earth Planet. Inter.* **115**, 1–16 (1999).
- Abercrombie, R. E. & Mori, J. J. Local observations of the onset of a large earthquakes: 28 June 1992 Landers, California. *Bull. Seismol. Soc. Am.* **84**, 725–734 (1994).
- Wesnousky, S. G. Seismological and structural evolution of strike-slip faults. *Nature* **335**, 340–343 (1988).
- Harris, R. A. & Day, S. M. Dynamic 3D simulations of earthquakes on en echelon faults. *Geophys. Res. Lett.* **26**, 2089–2092 (1999).

- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **2**, 297–356 (1981).
- Kanamori, H. & Anderson, D. L. Theoretical basis for some empirical relations in seismology. *Bull. Seismol. Soc. Am.* **65**, 1073–1095 (1975).
- Chen, Y. Thermal model of oceanic transform faults. *J. Geophys. Res.* **93**, 8839–8851 (1988).
- Detrick, R. S., White, R. S. & Purdy, G. M. Crustal structure of North Atlantic fracture zones. *Rev. Geophys.* **31**, 439–458 (1993).
- Brune, J. N. Seismic moment, seismicity and rate of slip along major fault zones. *J. Geophys. Res.* **73**, 777–784 (1968).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
- Smith, W. H. F. & Sandwell, D. T. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962 (1997).
- Wessel, P. & Smith, W. H. F. Free software helps map and display data. *Eos* **72**, 441, 445–446 (1991).
- Ekström, G. A very broadband inversion method for the recovery of earthquake source parameters. *Tectonophysics* **166**, 73–100 (1989).
- Antolik, M., Kaverina, A. & Dreger, D. Compound rupture of the great 1998 Antarctic plate earthquake. *J. Geophys. Res.* **105**, 23825–23838 (2000).
- McKenzie, D. P. Speculations on the consequences and causes of plate motions. *Geophys. J. R. Astron. Soc.* **18**, 1–32 (1969).

Acknowledgements

We thank M. Antolik for assistance with the slip inversion, and J. Tromp for discussion.

Correspondence and requests for materials should be addressed to R. E. A. (e-mail: rachel@seismology.harvard.edu).

Isotopic evidence for microbial sulphate reduction in the early Archaean era

Yanan Shen*, Roger Buick† & Donald E. Canfield*

* Danish Center for Earth System Science (DCESS) and Institute of Biology, Odense University, SDU, Campusvej 55, 5230 Odense M, Denmark

† School of Geosciences FO5, University of Sydney, Sydney, NSW 2006, Australia

Sulphate-reducing microbes affect the modern sulphur cycle, and may be quite ancient^{1,2}, though when they evolved is uncertain. These organisms produce sulphide while oxidizing organic matter or hydrogen with sulphate³. At sulphate concentrations greater than 1 mM, the sulphides are isotopically fractionated (depleted in ³⁴S) by 10–40‰ compared to the sulphate, with fractionations decreasing to near 0‰ at lower concentrations^{2,4–6}. The isotope record of sedimentary sulphides shows large fractionations relative to seawater sulphate by 2.7 Gyr ago, indicating microbial sulphate reduction⁷. In older rocks, however, much smaller fractionations are of equivocal origin, possibly biogenic but also possibly volcanogenic^{2,8–10}. Here we report microscopic sulphides in ~3.47-Gyr-old barites from North Pole, Australia, with maximum fractionations of 21.1‰, about a mean of 11.6‰, clearly indicating microbial sulphate reduction. Our results extend the geological record of microbial sulphate reduction back more than 750 million years, and represent direct evidence of an early specific metabolic pathway—allowing time calibration of a deep node on the tree of life.

Our samples came from the Dresser Formation (Warrawoona Group, Pilbara Craton) at North Pole in northwestern Australia. These rocks have experienced only very-low-grade metamorphism and slight deformation¹¹. Their age is constrained by zircon U–Pb geochronology to 3.515–3.458 Gyr (ref. 12). The Dresser Formation consists of pillowed, amygdaloidal basalt with three interbeds of cherty metasediments. Most samples came from the lowest chert bed, from lenses of bedded barite (BaSO₄) within silicified volcanogenic and carbonate sediments that were deposited in a shallow

subaqueous to intermittently exposed setting¹³. Other samples are from barite veins transecting and underlying the chert units. The barite beds are sedimentary deposits, indicated by crystals that were draped by detrital sediment (Fig. 1a), or were eroded and rounded during clastic redeposition. Measurements of interfacial angles show that the barite crystals were originally composed of gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$)^{11,14}, which precipitated in evaporating brine ponds separated from the sea by a permeable pumiceous sand bar¹⁵. Low sulphate concentrations in sea water seeping into the brine ponds may have been locally supplemented by the phototrophic oxidation of volcanogenic sulphide¹¹.

In much of the original gypsum, isomorphous substitution of Ba^{2+} for Ca^{2+} ('baritization') occurred soon after burial when relatively cool hydrothermal fluids circulated through the surrounding basalt pile¹¹. About 30% excess sulphate is required for isomorphous replacement because of differences in unit cell volume between parent gypsum and daughter barite. This extra sulphate was probably derived locally from the bedded gypsum where some of the original gypsum is now replaced by quartz¹¹, from dispersed sedimentary gypsum now also replaced by quartz¹¹, and from pore fluids. The hydrothermal fluids were an unlikely source of sulphate because the very low solubility of barite requires that Ba^{2+} and SO_4^{2-} cannot be transported together in solution. Sulphate remobilized from sedimentary gypsum could also have contributed to barite precipitation in the crosscutting veins. Thus, the barite veins may represent conduits for hydrothermal fluid circulation¹⁶, where

upwelling Ba^{2+} solutions mixed with surficial sulphate-containing solutions.

Between the barite beds are undulating laminae 2–10 mm thick composed of macroscopic pyrite (Fig. 1b), now usually weathered to iron oxides but preserved in subsurface samples. Within the barite are microscopic sulphide crystals $\sim 50 \mu\text{m}$ in size comprising less than 1% of the rock. These sulphides, mostly pyrite but some sphalerite, are aligned along growth faces of the original gypsum crystals (Fig. 1c). This attests to the syngenesis of sulphate and microscopic sulphide, and confirms that most sulphate is residual from the precursor gypsum. Also within the barite crystals are H_2S -containing fluid inclusions $\sim 10 \mu\text{m}$ in size, apparently formed during crystal growth¹⁷. Barite in the transecting veins contains fine, disseminated pyrite crystals, but lacks pyrite laminae and sulphureous fluid inclusions. Organic matter is finely dispersed throughout the bedded barite, but is absent from the vein barite.

Previous isotopic analyses of the macroscopic pyrite laminae revealed an average isotopic composition ($\delta^{34}\text{S}$; see Fig. 2 legend) of -0.9‰ (s.d. = 1.5)¹⁴, somewhat ^{34}S -depleted compared with the mantle value of 0‰ . Our data (Fig. 2; full analytical data available as Supplementary Information) are similar, though slightly more depleted in ^{34}S ($-2.4\text{‰} \pm 0.8$). While these pyrites could represent unfractionated volcanogenic sulphur, they are consistently more ^{34}S -depleted than is typical of younger volcanogenic sulphide deposits¹⁸. However, they are not sufficiently ^{34}S -depleted to be definitely biological, so their origin is still equivocal¹¹.

By contrast, the previously unanalysed microscopic sulphides are highly ^{34}S -depleted (Fig. 2), with fractionations relative to coexisting sulphate ($\epsilon_{\text{sulphide}}$) ranging from 21.1‰ to 7.4‰ , with a mean of 11.6‰ . Fractionations in the same range ($21.3\text{--}5.8\text{‰}$) are also produced if we calculate relative to the average isotopic composition of sulphate in the bedded barite. Non-biological processes can, in principle, produce $\epsilon_{\text{sulphide}}$ values of this magnitude. For example, hydrolysis of SO_2 to sulphate and sulphide occurs in relatively oxidizing magmatic fluids as temperatures drop below 400°C , forming H_2S depleted in ^{34}S by $15\text{--}20\text{‰}$ relative to sulphate¹⁸. However, the rocks surrounding the North Pole deposits are mafic¹¹ and would have generated reduced rather than oxidized magmatic fluids. Furthermore, ^{34}S -depleted sulphides precipitated with sulphates from Archaean magmatic fluids are usually associated with pervasive haematite deposition¹⁹, absent from the North Pole barite deposits. Therefore, the hydrolysis of SO_2 is an unlikely explanation for our results.

Alternatively, inorganic reduction of sulphate to sulphide by ferrous minerals can occur at near-neutral pH and temperatures greater than 200°C (ref. 18), with equilibrium $\epsilon_{\text{sulphide}}$ ranging from 27‰ at 200°C to $\sim 10\text{‰}$ at 550°C (ref. 18). A recirculating hydrothermal system with sulphate entrained from the brine ponds could conceivably have produced ^{34}S -depleted sulphides during partial sulphate reduction in hot underlying basalts. A spectrum of isotope values could, in principle, result from variable sulphate depletion or fluctuating temperatures during hydrothermal sulphate reduction. Our results, however, are largely outside the range, -5‰ to $+9\text{‰}$, generally found for high-temperature hydrothermal sulphate reduction⁸.

Furthermore, the original sulphate mineral was gypsum, not anhydrite or barite, the two typical hydrothermal sulphates. Gypsum is only stable below $\sim 60^\circ\text{C}$ (ref. 20), so the alignment of the microscopic sulphides along the crystal faces of the original gypsum (Fig. 1c) indicates that these sulphides were formed at relatively low temperatures along with the original gypsum. These microscopic sulphides were, therefore, formed before baritization, the earliest known hydrothermal event at North Pole¹³. We thus find no compelling evidence for a magmatic or hydrothermal origin for the microscopic sulphides in the bedded barites at North Pole. Given the early formation of the microscopic sulphides at low temperature, a metamorphic origin is also unlikely.

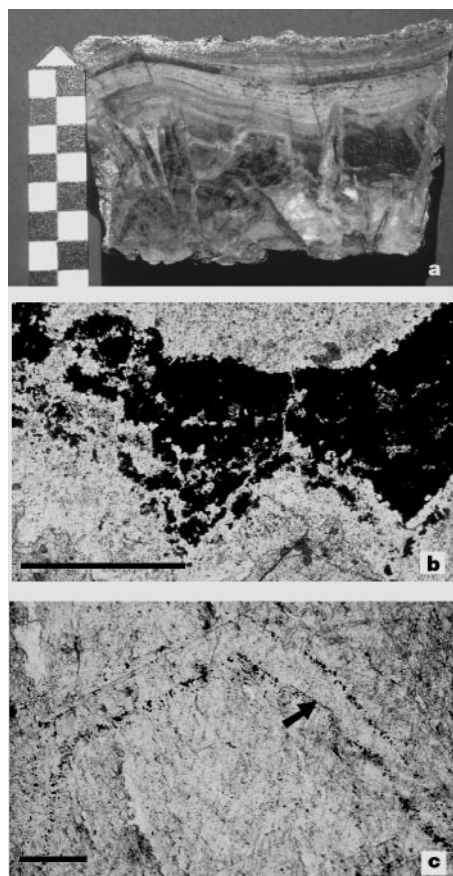


Figure 1 Relationships between sulphur species at North Pole, Australia. **a**, Laminae of silicified volcanoclastic/carbonate sediment draping over dark grey crystals of bedded barite, indicating a sedimentary origin for the precursor evaporative gypsum. The scale has 1-cm blocks, with an arrow pointing towards stratigraphic top. **b**, Opaque pyrite lamination (black) between barite beds draping over individual barite crystals. Scale bar, 1 mm. **c**, Microscopic sulphides (black, arrowed) included within a bedded crystal and aligned along growth faces of the precursor gypsum crystal. Scale bar, 1 mm.

The fractionations between sulphate and microcrystalline sulphides at North Pole are within the range observed for modern sulphate-reducing microbes metabolizing with > 1 mM sulphate, especially for organisms growing at high specific rates of sulphate reduction where ϵ_{SR} (fractionation during sulphate reduction) values of 10‰ to 25‰ are typical⁵. Moreover, the sulphides are intimately associated with organic carbon, the principal electron donor for biological sulphate reduction. These features, together with the lack of evidence for non-biological sources of highly fractionated sulphide, demonstrate that microbial sulphate reduction had evolved by 3.47 Gyr ago. Sulphate reduction is evident here because the North Pole evaporite ponds were localized oases of high sulphate concentrations, containing organic electron donors produced by an indigenous stromatolitic microbiota¹⁵, providing favourable conditions for sulphate-reducers to generate high fractionations. This contrasts with open marine Archaean settings, where only small fractionations of ~ 0 ‰ were expressed due to low sulphate concentrations and hence low atmospheric oxygen levels²¹.

Sulphides from the vein barites have a similar spread in $\epsilon_{sulphide}$ values as the bedded barites, ranging from 16.1‰ to 3.4‰. While the veins may have fed hydrothermal fluids to the sedimentary barite beds¹⁶, the vein sulphides show no evidence for a magmatic, hydrothermal or metamorphic origin. It is more likely that these sulphides were biogenic, although it is unclear whether they

were formed *in situ* or remobilized from surficial environments. However, the absence of sulphureous fluid inclusions in the vein barite suggests the latter.

Our results provide the oldest evidence of microbial sulphate reduction in the geological record, pre-dating previous evidence⁷ by more than 0.75 Gyr (Fig. 3). They also give the earliest indication of a specific microbial metabolism. Though the isotopic record of organic carbon suggests that autotrophic CO_2 fixation into biomass had evolved by ~ 3.8 Gyr ago^{22,23}, the specific metabolic pathways employed, and the organisms responsible, are unclear. Likewise, although microfossil evidence is permissive of cyanobacteria by 3.45 Gyr ago²⁴, these simple forms are not diagnostic of a specific microbial affinity or a particular metabolic pathway⁶. Sulphate reduction is a complex metabolic process requiring advanced membrane-bound transport enzymes, proton motive force generation by ATPase and other charge separation proteins, and the genetic regulation of protein synthesis through DNA and RNA²⁵. Therefore, by 3.47 Gyr ago—and probably earlier as the known pathways of CO_2 fixation are also complex—microbes had already developed many of the critical cellular systems shared by their modern descendants.

Dissimilatory sulphate reduction occurs in both the Archaeal and Bacterial domains of the tree of life as deduced from small-subunit ribosomal RNA (SSU rRNA). Among the Archaea, the only known

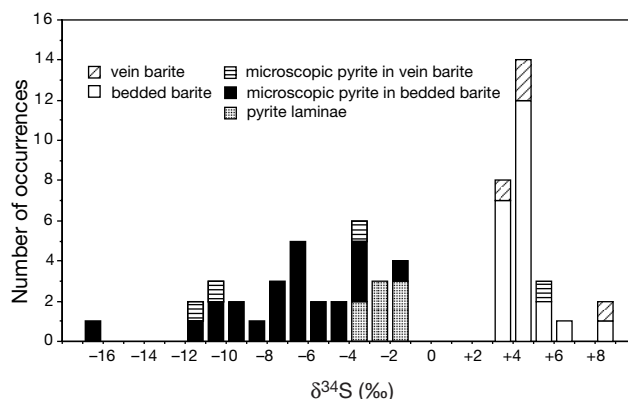


Figure 2 The isotopic composition of sulphur species from the barite deposits of North Pole, Australia. $\delta^{34}S = 1,000 \{ [(^{34}S/^{32}S)_{\text{sample}} / (^{34}S/^{32}S)_{\text{standard}}] - 1 \}$ relative to the Cañon Diablo Troilite standard.

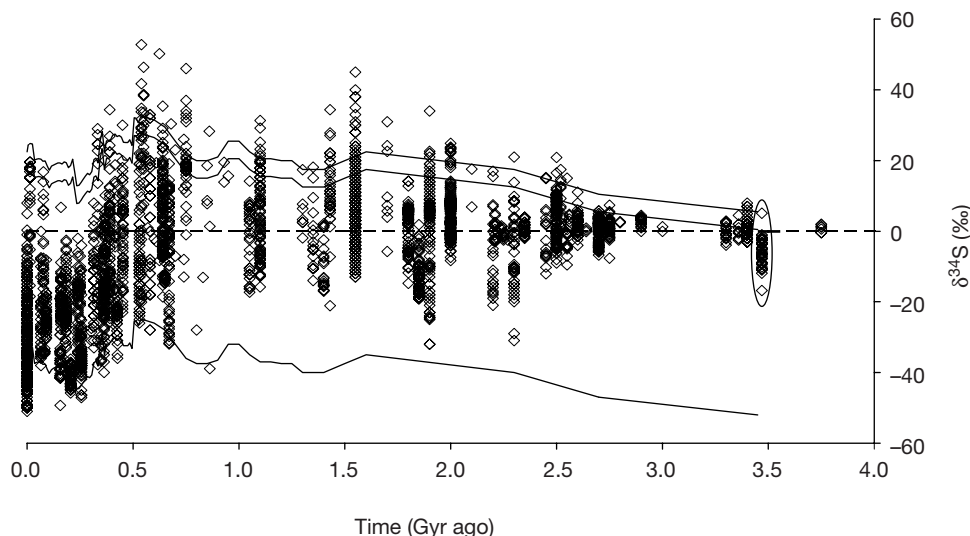


Figure 3 The secular trends in the isotopic composition of seawater sulphate and sulphide over geological time. Data within the oval are from this work, the other data are from refs 2, 29. The band (double line) in the upper part of the figure represents the isotopic composition of seawater sulphate through time. The single line in the lower part of

the figure is displaced from the seawater sulphate trend by 55‰, representing the maximum fractionation between sulphate and sulphide through the past 600 million years. Before 1.7 Gyr ago, constraints on the isotopic composition of seawater sulphate are sparse.

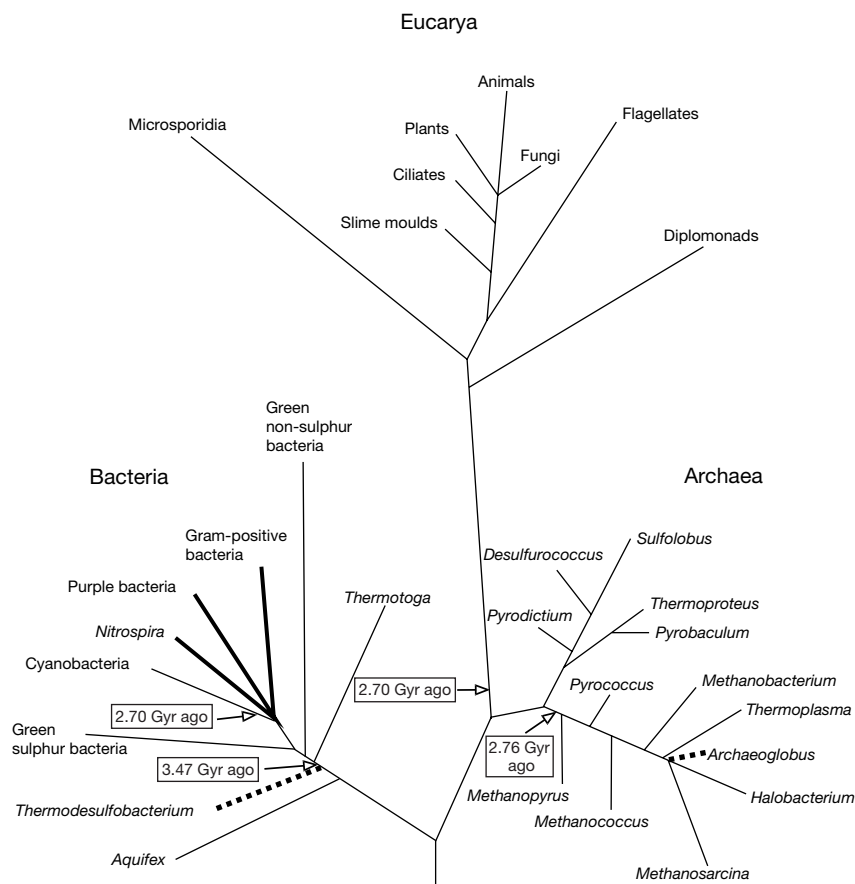


Figure 4 The tree of life based on SSU rRNA sequence analysis, with some temporal constraints on branching. The tree is modified from ref. 2, and abstracted from phylogenetic trees presented in refs 26 and 27. The time calibration points are from ref. 30, with our additional constraint of 3.47 Gyr placed in the Bacterial domain. Lineages

housing sulphate-reducers metabolizing at temperatures $> 70^{\circ}\text{C}$ are shown by broken black lines, while lineages supporting sulphate-reducers metabolizing at $< 70^{\circ}\text{C}$ are shown by heavy black lines.

sulphate-reducers are hyperthermophiles with optimal growth temperature above 80°C , and these are restricted to the single genus *Archaeoglobus*²⁶. Within the Bacteria, the most deeply branching sulphate-reducers belong to the genus *Thermodesulfobacterium*^{2,27}, which are also hyperthermophiles with optimal growth around 80°C . Thus far, sulphate reduction by organisms metabolizing at temperatures below 70°C is known only from the δ -subdivision of the proteobacteria (purple bacteria), the Gram-positive bacteria, and the Nitrospira group (Fig. 4). As only a small proportion of total microbial diversity has been characterized²⁷, it is possible that the ability to reduce sulphate at moderate temperatures resides elsewhere in prokaryote phylogeny.

As the original gypsum now forming the bedded barites at North Pole first precipitated at temperatures below 60°C , the organisms responsible for sulphate reduction were apparently mesophiles or, at most, moderate thermophiles. Given what is currently known about the phylogenetic distribution of temperature adaptations among sulphate-reducers, our findings suggests a minimum age of 3.47 Gyr for a node immediately above the branching point of the hyperthermophilic *Thermodesulfobacterium* lineage in the Bacterial domain (Fig. 4). This placement is necessarily tentative, as deeper-branching mesophilic sulphate-reducers may be discovered, but even so it represents the oldest evolutionary event thus far dated on the tree of life. □

Methods

Macroscopic pyrites were sampled with a dental drill, microscopic sulphides were distilled from the enclosing barite by Cr-reduction²⁸ and collected as Ag_2S , and residual barite was collected after Cr-reduction. Isotope analysis was performed in two ways. For large

samples, SO_2 gas was generated from both pyrite and barite by heating the sulphur samples to $1,050^{\circ}\text{C}$ and $1,150^{\circ}\text{C}$, respectively, with cuprous oxide. The SO_2 gas was purified on a high-vacuum gas extraction line and subsequent isotope analyses were reproducible to better than $\pm 0.5\text{‰}$. Small samples of barite and Ag_2S were combusted directly to SO_2 in an elemental analyser coupled directly to an isotope ratio mass spectrometer (isotope analyses reproducible to $\pm 0.3\text{‰}$). Comparison of both techniques on homogenized splits of the same samples showed an average difference of 0.24‰ ($n = 5$).

Received 18 May; accepted 24 November 2000.

- Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A. & Stahl, D. A. Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* **180**, 2975–2982 (1998).
- Canfield, D. E. & Raiswell, R. The evolution of the sulfur cycle. *Am. J. Sci.* **299**, 697–723 (1999).
- Postgate, J. R. *The Sulphate-Reducing Bacteria* 2nd edn (Cambridge Univ. Press, Cambridge, 1984).
- Harrison, A. G. & Thode, H. G. Mechanisms of the bacterial reduction of sulphate from isotope fractionation studies. *Trans. Faraday Soc.* **53**, 84–92 (1958).
- Kaplan, I. R. & Rittenberg, S. C. Microbiological fractionation of sulphur isotopes. *J. Gen. Microbiol.* **34**, 195–212 (1964).
- Knoll, A. H. & Canfield, D. E. Isotopic inferences on early ecosystems. *Paleontol. Soc. Pap.* **4**, 212–243 (1998).
- Goodwin, A. M., Monster, J. & Thode, H. G. Carbon and sulfur isotope abundances in Archean iron formations and early Precambrian life. *Econ. Geol.* **71**, 870–891 (1976).
- Cameron, E. M. Sulphate and sulphate reduction in early Precambrian. *Nature* **296**, 145–148 (1982).
- Hayes, J. M., Lambert, I. B. & Strauss, H. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 129–134 (Cambridge Univ. Press, Cambridge, 1992).
- Ohmoto, H., Kakegawa, T. & Lowe, D. R. 3.4-billion-year-old biogenic pyrites from Barberton, South Africa: sulfur isotope evidence. *Science* **262**, 555–557 (1993).
- Buick, R. & Dunlop, J. S. R. Evaporitic sediments of early Archean age from the Warrawoona Group, North Pole, Western Australia. *Sedimentology* **37**, 247–277 (1990).
- Buick, R. *et al.* Record of emergent continental crust ~3.5 billion years ago in the Pilbara Craton of Australia. *Nature* **375**, 574–577 (1995).
- Buick, R. & Barnes, K. R. Cherts in the Warrawoona Group: early Archean silicified sediments deposited in shallow water environments. *Univ. West. Aust. Geol. Dept. Univ. Extension Spec. Publ.* **9**, 37–53 (1984).
- Lambert, I. B., Donnelly, T. H., Dunlop, J. S. R. & Groves, D. I. Stable isotope compositions of early Archean sulphate deposits of probable evaporite and volcanogenic origins. *Nature* **276**, 808–810 (1978).

15. Groves, D. I., Dunlop, J. S. R. & Buick, R. An early habitat of life. *Sci. Am.* **245**, 64–73 (1981).
16. Nijman, W., de Bruijne, K. C. H. & Valkering, M. E. Growth fault control of Early Archaean cherts, barite mounds and chert-barite veins, North Pole Dome, Eastern Pilbara, Western Australia. *Precamb. Res.* **88**, 25–52 (1999).
17. Rankin, A. H. & Shepherd, T. J. H₂S-bearing fluid inclusions in baryte from the North Pole deposit, Western Australia. *Mineral. Mag.* **42**, 408–410 (1978).
18. Ohmoto, H. & Goldhaber, M. B. in *Geochemistry of Hydrothermal Ore Deposits* (ed. Barnes, H. L.) 517–611 (Wiley, New York, 1997).
19. Cameron, E. M. & Hattori, K. Archaean gold mineralization and oxidized hydrothermal fluids. *Econ. Geol.* **82**, 1177–1191 (1987).
20. Hardie, L. A. The gypsum-anhydrite equilibrium at one atmosphere pressure. *Am. Mineral.* **52**, 171–200 (1967).
21. Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661 (2000).
22. Schidlowski, M. A 3800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature* **333**, 313–318 (1988).
23. Rosing, M. T. ¹³C-depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from west Greenland. *Science* **283**, 674–676 (1999).
24. Schopf, J. W. & Packer, B. M. Early Archaean (3.3-billion to 3.5-billion-year-old) microfossils from the Warrawoona Group, Australia. *Science* **237**, 70–73 (1987).
25. Widdel, F. in *Biology of Anaerobic Microorganisms* (ed. Zehnder, A. J. B.) 496–585 (Wiley, New York, 1988).
26. Stetter, K. O. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)* (eds Bock, G. R. & Goode, J. A.) 1–10 (Wiley, New York, 1996).
27. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
28. Canfield, D. E., Raiswell, R., Westrich, J. T., Reaves, C. M. & Berner, R. A. The use of chromium reduction in the analysis of reduced inorganic sulfur in sediments and shales. *Chem. Geol.* **54**, 149–155 (1986).
29. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
30. Knoll, A. H. A new molecular window on early life. *Science* **285**, 1025–1026 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. S. R. Dunlop for suggesting that we should examine the isotopic systematics of microcosmic sulphur species in the North Pole barite; K.-U. Hinrichs, K. Londry, R. Summons, B. Thamdrup and K. Habicht for discussions; I. O'Brien, O. Thomas and L. Salling for technical assistance; and D. Des Marais for comments and suggestions. This work was supported by the Danish Grundforskningsfond (Basic Research Foundation) and by the Australian Research Council (R.B.).

Correspondence and requests for materials should be addressed to Y.S. (e-mail: shen@biology.ou.dk).

A primitive sarcopterygian fish with an eyestalk

Min Zhu*, Xiaobo Yu† & Per E. Ahlberg‡

* Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China

† Department of Biological Sciences, Kean University, Union, New Jersey 07083, USA

‡ Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

The discovery of two Early Devonian osteichthyan (bony fish) fossils^{1–4} has challenged established ideas about the origin of osteichthyans and their divergence into actinopterygians (teleosts and their relatives) and sarcopterygians (tetrapods, coelacanths, lungfishes and related groups)^{5–7}. *Psarolepis* from China^{1,2,8,9} and an unnamed braincase from Australia³ combine derived sarcopterygian and actinopterygian characters with primitive features previously restricted to non-osteichthyans, suggesting that early osteichthyan evolution may have involved substantial parallelism between sarcopterygians and actinopterygians. But interpretation of these fossils has been hampered by poor phylogenetic resolution^{1,3}. Here we describe a basal sarcopterygian fish, *Achoania* gen. et sp. nov., that fills the morphological gap between *Psarolepis* and higher sarcopterygians. We also report the

presence of eyestalk attachments in both *Achoania* and *Psarolepis*, showing that this supposedly non-osteichthyan feature occurs in basal sarcopterygians as well as the actinopterygian-like Australian braincase³.

Sarcopterygii (Romer, 1955)

Achoania gen. nov.

Diagnosis. A sarcopterygian with an anteroventrally sloping intra-cranial joint; robust and rod-like premaxilla that lacks a posterodorsal process and does not contribute to the orbital margin; toothed median rostral that does not separate the premaxillae; large internasal cavities occupying about 33% of the length of the ethmosphenoid floor; small drop-shaped parasphenoid flanked by flat roughened surface; robust postorbital pila; large descending processes on ventral surface of sphenoidal region; very wide sub-orbital ledge; and prominent eyestalk scar. The external bones are covered by large-pore cosmine similar to that in *Psarolepis*.

Type species. *Achoania jarvikii* sp. nov.

Diagnosis. As for genus.

Etymology. Generic name referring to the absence of choana, from Greek *a* (not) and *choane* (funnel). Specific name in honour of the late Erik Jarvik.

Holotype. V6235, a fairly complete anterior cranial portion, IVPP, Beijing.

Locality and horizon. Qujing, E. Yunnan, China. Xitun Formation (late Lochkovian, Early Devonian)

Remarks. The new genus is represented by one anterior cranial portion (Fig. 1), which differs substantially from *Psarolepis* in the shape of the premaxilla, the parasphenoid, the internasal cavity (which occupies more than 70% of the length of the ethmosphenoid floor in *Psarolepis*), the postorbital pila and structures on the ventral side of the sphenoid. All the anterior cranial portions of *Psarolepis* are uniformly distinct from *Achoania* with no intermediate variations, and the posterior cranial specimens and lower jaws of *Psarolepis* fail to match the morphology of *Achoania*.

Description. The most impressive feature of *Achoania* gen. et sp. nov., is a heart-shaped unfinished area (Fig. 1c, d) in the interorbital wall, which lies immediately behind the optic canal and has outward-facing edges. Two small cup-shaped recesses lie posterodorsal and ventral to this area, the posterodorsal recess supported by a posteriorly running horizontal ridge. By comparison with placoderms^{7,10}, primitive chondrichthyans^{7,11} and the Australian braincase (AMF101607)^{3,4}, the unfinished area represents the eyestalk attachment area, and the two small recesses represent sites for eye muscle attachment.

Except for its recently reported presence in AMF101607 (ref. 3), an eyestalk had been considered to exist only in two non-osteichthyan groups, chondrichthyans and the extinct placoderms. The presence of an eyestalk area in *Achoania* is corroborated by the discovery of a similar structure in a newly collected *Psarolepis* specimen (Fig. 2), which shows, on both sides of the interorbital wall, a large unfinished area with well-defined margins between the optic nerve canal and pituitary vein canal. In addition, re-examination of a previously described specimen (V8136)⁹ reveals a similarly positioned, although distorted, unfinished area on the exposed side of the interorbital wall. The well-preserved eyestalk area on both sides of the new *Psarolepis* specimen and the similar arrangement of foramina, pits and ridges in the interorbital wall of *Psarolepis* and *Achoania* strongly indicate that the eyestalk area in these two forms is a natural structure rather than an artefact of preservation or preparation. The eyestalk should now be considered a general gnathostome feature retained by early osteichthyans on both the actinopterygian and sarcopterygian lineages.

In other cranial features, *Achoania* resembles *Psarolepis* and differs from previously known sarcopterygians in having large, closely spaced pores on the cosmine surface, an anterodorsally facing anterior nostril, a toothed median rostral, a straight rather than lyre-shaped trajectory of the supraorbital sensory canal and a

short parasphenoid not pierced by internal carotid canals. Like *Psarolepis*⁹, *Achoania* has a postorbital pila which forms a bridge connecting the top of the basiptyergoid process and the side of the braincase wall, and laterally flanks the canal for the jugular vein.

Achoania differs from *Psarolepis* and agrees with typical sarcopterygians in having a premaxilla without a posterodorsal process. Well-preserved sutural surfaces indicate that the posterior nostril was enclosed posteriorly by a separate bone, presumably the lacrimal as in other sarcopterygians. The sphenoidal region of *Achoania* is longer than in *Psarolepis*, but shorter than in other osteichthyans (except *Onychodus*). *Achoania* thus bridges the morphological gap between *Psarolepis* and other sarcopterygians, and helps to resolve their relationships.

A phylogenetic analysis, based on a new data matrix of 158 morphological characters, places *Achoania* and *Psarolepis* as successive sister groups to all other sarcopterygians (Fig. 3a). Clade decay values for surrounding nodes are relatively high, indicating that their position is well supported. The position of the Australian braincase AMF101607 is less clear, but it is probably a very primitive actinopterygian. This topology matches one of the alternative topologies that were put forward (but could not be resolved) in previous analyses^{1–3}.

The new phylogeny has major evolutionary and biogeographical implications. Past analyses of osteichthyans^{5,6,12,13} have generated inferred ancestral character complements for the clade based on known 'primitive' genera such as *Mimia* and *Porolepis*, but these differ substantially from the character suite revealed by *Achoania*, *Psarolepis* and AMF101607. Our well-supported resolution of *Achoania* and *Psarolepis* as stem-group sarcopterygians shows that their seemingly non-osteichthyan characters, such as an eyestalk, a placoderm-like shoulder girdle with a separate spinal plate, and paired and median fin spines^{1,8}, are present in the base of the osteichthyan crown group; in all likelihood these characters are primitive for the Osteichthyes as a whole. This implies that many

apparent synapomorphies between previously known actinopterygians and sarcopterygians (undivided cleithrum, lack of eyestalk, lack of fin spines) are convergent between the two lineages. We can no longer assume that direct comparisons between derived sarcopterygians and actinopterygians will be informative about primitive conditions for the Osteichthyes as a whole. Instead, attention should focus on basal taxa, such as the Early Devonian ?actinopterygian

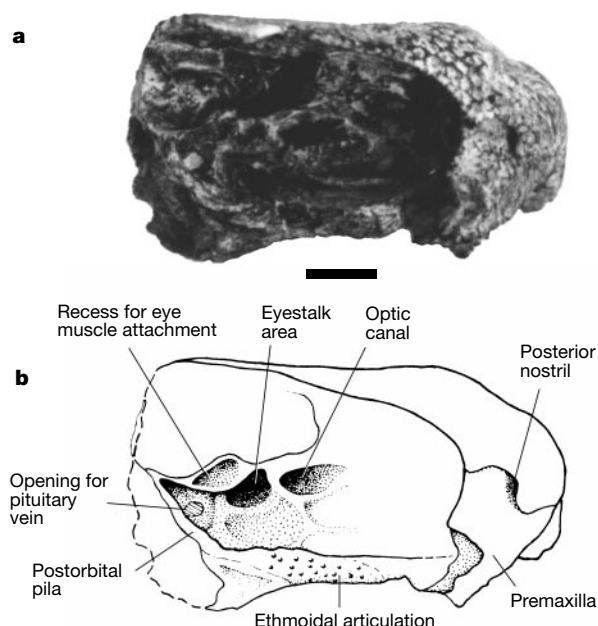


Figure 2 *Psarolepis* specimen (IVPP V11490.1) in posterolateral view, showing the eyestalk attachment area. **a**, Photograph; **b**, drawing. Scale bar, 2 mm.

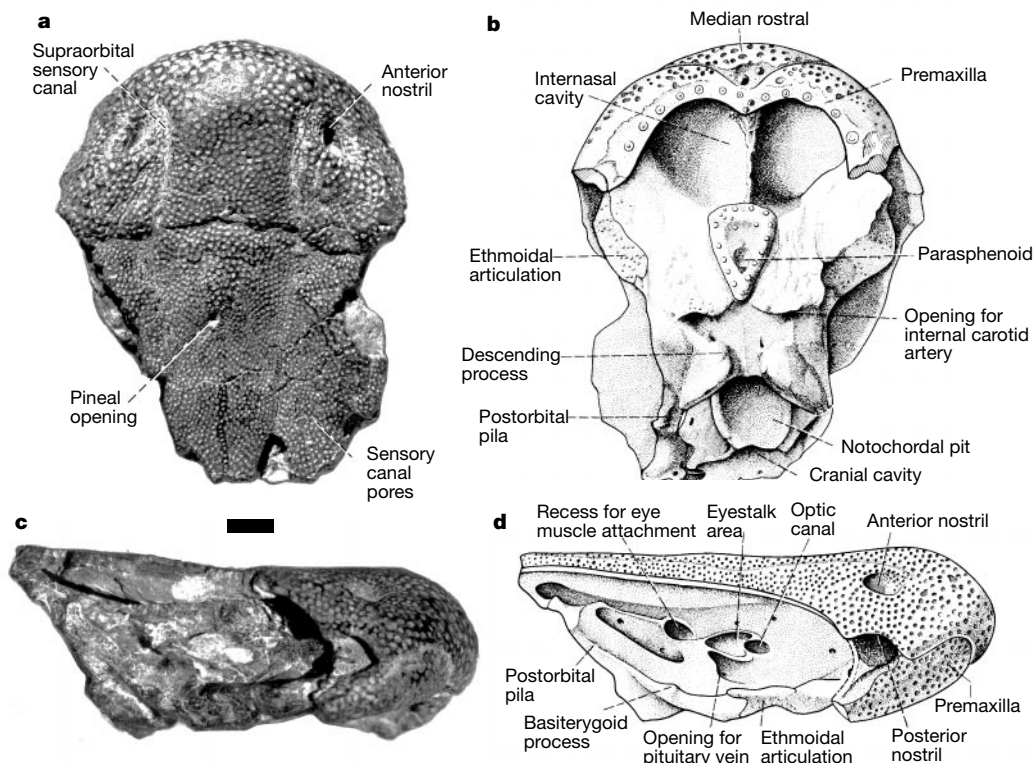


Figure 1 *Achoania jarviki* gen. et sp. nov., holotype, IVPP V6835. **a**, Dorsal view (photograph), showing large-pore cosmine similar to that in *Psarolepis*. **b**, Ventral view (drawing), showing toothed median rostral, large internasal

cavities, small drop-shaped parasphenoid. **c**, **d**, Right lateral views of same specimen showing the eyestalk attachment area. Scale bar, 2 mm.

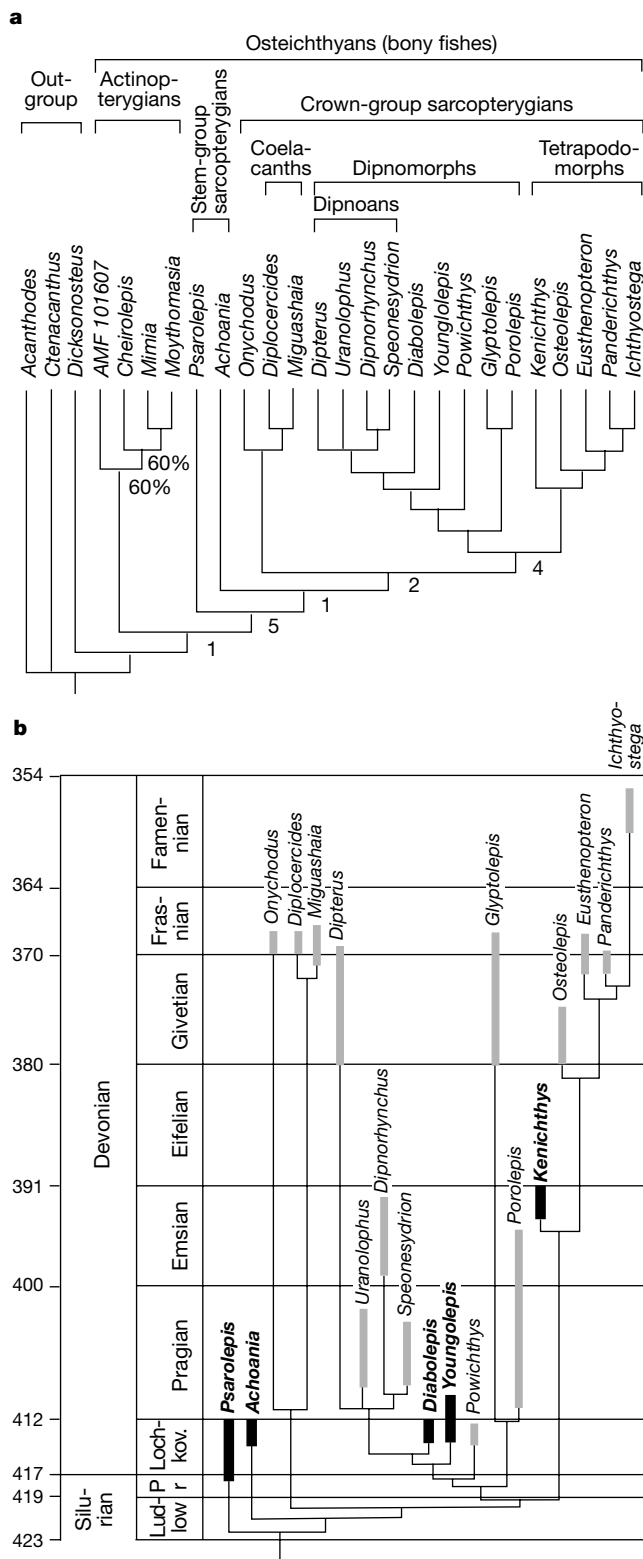


Figure 3 Phylogenetic analysis. **a**, Fifty per cent majority rule consensus tree calculated from 15 equally parsimonious trees of length 317 steps; consistency index 0.568, homoplasy index 0.432, retention index 0.765, and rescaled consistency index 0.434. All branches have 100% support except those that define the position of AMF101607, which have only 60% support. The topology is the same as the strict consensus tree except that the nodes with 60% support collapse in the latter. Numbers below the tree indicate the Bremer Support/clade decay values for the adjacent nodes. **b**, The sarcopterygian part of the tree projected against stratigraphy. Lochkovian, Lochkovian; Pr, Pridoli. Taxa from Yunnan are shown as broad black range bars with names in bold print. Dates along left margin in millions of years.

*Dialipina*¹⁴ which has a shoulder girdle comparable in structure to that of *Psarolepis*. We must also reassess some supposed autapomorphies of different osteichthyan groups, such as the 'extracleithrum' of coelacanth which, in the primitive genus *Miguashaia*^{15,16}, closely resembles a spinal plate.

The character distributions observed in *Achoania*, *Psarolepis* and AMF101607 make basal Osteichthyes appear more placoderm-like than recognized previously. Many workers place the placoderms as the most primitive jawed gnathostomes⁷ and explicitly deny any homology between the placoderm and osteichthyan bone patterns, thus virtually guaranteeing wide phylogenetic separation. However, evidence is mounting for homology between the dermal shoulder girdles of osteichthyan and placoderms. Our analysis places the placoderm *Dicksonosteus* as the sister group of Osteichthyes, although the branching pattern within the Osteichthyes is unaffected by the removal of *Dicksonosteus* (or any other taxon) from the outgroup. Deep gnathostome phylogeny needs to be reviewed in the light of these findings.

The occurrence of two stem sarcopterygians (*Psarolepis* and *Achoania*) and three basal dipnomorphs (*Powichthys*¹⁷, *Youngolepis*¹⁸ and *Diabolepis*¹⁹) in the Lochkovian, but only one (*Psarolepis*) in the underlying Pridoli, suggests that the Sarcopterygii were diversifying rapidly around the Silurian/Devonian boundary. Stratigraphic mapping of the 50% majority rule tree (Fig. 3b) assigns a very brief duration to the internode between coelacanth and lungfishes plus tetrapods. This may help to explain the difficulty of resolving the coelacanth–lungfish–tetrapod three-taxon problem using molecular data^{20,21}.

The discovery of *Achoania* highlights the importance of Yunnan as a source of early sarcopterygians. Yunnan has yielded the whole sarcopterygian stem group (*Psarolepis* and *Achoania*), the earliest and most basal tetrapodomorph (*Kenichthys*), and two of the four most basal dipnomorphs (*Youngolepis* and *Diabolepis*). Five of the six described Lochkovian (basal Devonian) sarcopterygians come from the south China terrane—a palaeocontinent encompassing Yunnan and northern Vietnam which lay adjacent to eastern Gondwana during the Early Devonian—and four of these (*Psarolepis*, *Achoania*, *Youngolepis* and *Diabolepis*) occur in the Xitun Formation of Yunnan^{9,18,19,22}. Further sarcopterygian material from the Xitun formation awaits description.

The Early Devonian vertebrate fauna of the south China terrane is highly endemic, and the reasonably abundant Lochkovian faunas from other areas almost always lack sarcopterygians and other south China elements²². This suggests that south China may be the place of origin of the Sarcopterygii. Future work on deep osteichthyan phylogeny should aim to test this hypothesis. □

Methods

Phylogenetic analysis

The analysis was performed by using the phylogenetic package PAUP 3.1 (ref. 24) with a data matrix of 26 taxa and 157 morphological characters (see Supplementary Information). Most parsimonious trees were obtained heuristically, using 500 replicates of random stepwise addition under the tree bisection–reconnection algorithm. All characters were unordered and unweighted.

Received 22 August; accepted 8 November 2000.

- Zhu, M., Yu, X. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397**, 607–610 (1999).
- Ahlberg, P. E. Something fishy in the family tree. *Nature* **397**, 564–565 (1999).
- Basden, A. M. *et al.* The most primitive osteichthyan braincase? *Nature* **403**, 185–188 (2000).
- Chang, M. M. Fossil fish up for election. *Nature* **403**, 152–153 (2000).
- Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Myrthomasia* from the Upper Devonian of Western Australia. *Bull. Br. Mus. Nat. Hist.* **37**, 173–428 (1984).
- Cloutier, R. & Ahlberg, P. E. in *Interrelationships of Fishes* (eds Stiassny, M. L. J., Parenti, L. R. & Johnson, G. D.) 445–479 (Academic, San Diego, 1996).
- Janvier, P. *Early Vertebrates* (Oxford Univ. Press, Oxford, 1996).
- Zhu, M. & Schultze, H.-P. The oldest sarcopterygian fish. *Lethaia* **30**, 293–304 (1997).
- Yu, X. A new porolepiform-like fish, *Psarolepis romeri*, gen. et. sp. nov. (Sarcopterygii, Osteichthyes) from the Lower Devonian of Yunnan, China. *J. Vert. Paleont.* **18**, 261–274 (1998).

10. Goujet, D. *Les Poissons Placodermes du Spitsberg. Arthroires Dolichothoraci de la Formation de Wood Bay (Dévonien Inférieur)* (Cahiers de Paléontologie, Section Vertébrés, Paris, 1984).
11. Maissey, J. G. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) Ch. 16 (Taylor & Francis, London, in press).
12. Ahlberg, P. E. A re-examination of sarcopterygian interrelationships, with special reference to the Porolepiformes. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).
13. Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. Lungfishes, tetrapods, paleontology and plesiomorphy. *Bull. Am. Mus. Nat. Hist.* **167**, 159–276 (1981).
14. Schultze, H.-P. & Cumbaa, S. L. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) Ch. 18 (Taylor & Francis, London, in press).
15. Cloutier, R. in *Devonian Fishes and Plants of Miguasha, Quebec, Canada* (eds Schultze, H.-P. & Cloutier, R.) 227–247 (Dr Friedrich Pfeil, Munich, 1996).
16. Forey, P. L. et al. A new coelacanth from the Middle Devonian of Latvia. *J. Vert. Paleont.* **20**, 243–252 (2000).
17. Jessen, H. L. Lower Devonian Porolepiformes from the Canadian Arctic with special reference to *Powichthys thorsteinsoni* Jessen. *Palaeontographica A* **167**, 180–214 (1980).
18. Chang, M. M. *The braincase of Youngolepis, a Lower Devonian Glossopterygian from Yunnan, south-western China*. Thesis, Stockholm Univ. (1982).
19. Chang, M. M. *Diabolepis* and its bearing on the relationships between porolepiforms and dipnoans. *Bull. Mus. Natl. Hist. Nat.* **17**, 235–268 (1995).
20. Zardoya, R. & Meyer, A. Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. *Proc. Natl. Acad. Sci. USA* **93**, 5449–5454 (1996).
21. Zardoya, R. et al. Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. *Mol. Biol. Evol.* **15**, 506–517 (1998).
22. Janvier, P. & Phuong, T. H. Les vertébrés (Placodermi, Galeaspidi) du Dévonien inférieur de la coupe de Lung Cò—Mia Lé, Province de Hà Giang, Viêt Nam, avec des données complémentaires sur les gisements à vertébrés du Dévonien inférieur du Bac Bo oriental. *Geodiversitas* **21**, 33–67 (1999).
23. Long, J. A. (ed.) *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (Belhaven, London, 1993).
24. Swofford, D. L. *PAUP: phylogenetic analysis using parsimony, ver.3.1.1* (Illinois Natural History Survey, Campaigne, Illinois, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. M. Chang, P. Janvier and M. I. Coates for useful discussions, U. Samuelson and Zhang Jie for photographic work, Lu Xiufen for specimen preparation and Hu Huiqing for artwork. M.Z. & X.Y. acknowledge the support from the Chinese Foundation of Natural Sciences, Ministry of Science & Technology (China), and National Geographic Society (US). X.Y. thanks Kean University for faculty research and development support.

Correspondence and requests for materials should be addressed to M.Z.

Modulation of the neuronal glutamate transporter EAAC1 by the interacting protein GTRAP3-18

Chien-liang Glenn Lin*, Irina Orlov*, Alicia M. Ruggiero, Margaret Dykes-Hoberg, Andy Lee, Mandy Jackson & Jeffrey D. Rothstein

Johns Hopkins University, Department of Neurology and Neuroscience, Meyer 6-109, 600 North Wolfe Street, Baltimore, Maryland 21287, USA

* These authors contributed equally to this work

Excitatory amino-acid carrier 1 (EAAC1) is a high-affinity Na^+ -dependent L-glutamate/D, L-aspartate cell-membrane transport protein¹. It is expressed in brain as well as several non-nervous tissues. In brain, EAAC1 is the primary neuronal glutamate transporter^{2,3}. It has a polarized distribution in cells and mainly functions perisynaptically to transport glutamate from the extracellular environment^{2–4}. In the kidney it is involved in renal acidic amino-acid re-absorption and amino-acid metabolism^{5–7}. Here we describe the identification and characterization of an EAAC1-associated protein, GTRAP3-18. Like EAAC1, GTRAP3-18 is expressed in numerous tissues^{8,9}. It localizes to the cell membrane and cytoplasm, and specifically interacts with carboxy-terminal intracellular domain of EAAC1. Increasing the expression of GTRAP3-18 in cells reduces EAAC1-mediated glutamate transport by lowering substrate affinity. The expression of GTRAP3-18

can be upregulated by retinoic acid, which results in a specific reduction of EAAC1-mediated glutamate transport. These studies show that glutamate transport proteins can be regulated potently and that GTRAP3-18 may be important in regulating the metabolic function of EAAC1.

Using the C-terminal intracellular domain of EAAC1 (the last 87 amino acids) as bait in a yeast two-hybrid screen of an adult rat brain complementary DNA library, we isolated 78 clones displaying β -galactosidase activity. Restriction and sequencing analyses revealed that 10 of the clones with the strongest β -galactosidase activity were identical. This clone, designated E18, was completely sequenced and was found to be unique after GenBank analysis. JWA protein, (GenBank NP006398) a human, differentially displayed, vitamin-A-responsive gene, is 95% identical to E18, suggesting that E18 is a JWA protein homologue of rat.

E18 is a full-length complementary DNA containing an initiation methionine and a poly(A) tail. We named the protein glutamate transporter EAAC1-associated protein (GTRAP3-18). GTRAP3-18 encodes a protein of 188 amino acids (see Supplementary Information), with a calculated relative molecular mass of 22,500 (M_r , 22.5K). Protein analysis indicated that it is a very hydrophobic protein with four possible transmembrane domains. Both the C-terminal and amino-terminal domains contain protein kinase C motifs and may be intracellular.

To confirm the yeast two-hybrid results, we examined the interaction of GTRAP3-18 with EAAC1 using *in vitro* and *in vivo* methods. For *in vitro* cell-free binding, EAAC1 was expressed as a fusion protein with glutathione S-transferase (GST), and GTRAP3-18 was produced and labelled with [³⁵S]methionine by *in vitro* transcription and translation. Purified GST or GST–EAAC1 fusion proteins immobilized on glutathione–sepharose were incubated with [³⁵S]labelled GTRAP3-18 protein. GTRAP3-18 bound specifically to immobilized GST–EAAC1 (lane 2) but not to GST alone (lane 3), indicating that they interact *in vitro* (Fig. 1a).

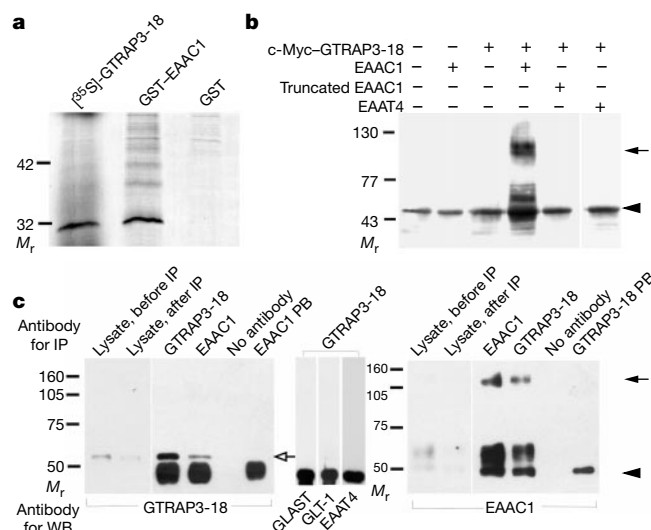


Figure 1 GTRAP3-18 interacts with EAAC1 *in vitro* and *in vivo*. **a**, SDS–PAGE analysis of cell-free *in vitro* binding. *In vitro* synthesized [³⁵S]-labelled GTRAP3-18 bound specifically to immobilized GST–EAAC1 and not to the negative control GST. **b**, Western blot analyses of immunoprecipitations from transfected HEK293 cells extracts, using anti-c-Myc antibodies for immunoprecipitation and anti-EAAC1 or EAAT4 antibodies for western blot. EAAC1 (monomer, arrowhead; dimer, arrow) was specifically co-immunoprecipitated with c-Myc–GTRAP3-18. An immunoprecipitation artifact is present in each lane at about 50–60K. **c**, Western blot (WB) analyses of immunoprecipitations (IP) from rat brain extracts. GTRAP3-18 (open arrowhead) was specifically co-immunoprecipitated with EAAC1. EAAC1 (monomer, arrowhead; dimer, arrow) was co-immunoprecipitated with GTRAP3-18. PB, peptide block.

We carried out immunoprecipitation experiments to test whether EAAC1 and GTRAP3-18 interact *in vivo*. Initial experiments were performed in transfected HEK293 cells using N-terminal c-Myc-tagged GTRAP3-18. As shown in Fig. 1b, EAAC1 co-immunoprecipitated with c-Myc-GTRAP3-18 in the cell extract prepared from co-expression cells (lane 4), but not from EAAC1 (lane 2) or c-Myc-GTRAP3-18 (lane 3) single-expression cells, which eliminated the possibility of artefacts arising from nonspecific immunobead binding or an antibody cross-reaction, respectively. A truncated EAAC1 lacking the interacting C-terminal domain (see below) did not co-immunoprecipitate with c-Myc-GTRAP3-18 (lane 5), further indicating an interaction between EAAC1 and GTRAP3-18. This interaction was specific, as EAAT4, another neuronal glutamate transporter subtype, was not immunoprecipitated with c-Myc-GTRAP3-18 (lane 6). Identical results were obtained using other cell lines (COS-7 and C6 glioma; data not shown).

To study the protein interaction *in vivo*, we used anti-EAAC1 or GTRAP3-18 polyclonal antibodies to immunoprecipitate EAAC1 or GTRAP3-18 from rat brain extract. EAAC1 co-immunoprecipitated specifically with GTRAP3-18, but not GLAST, GLT-1 or EAAT4 (Fig. 1c). Similarly, GTRAP3-18 co-immunoprecipitated with EAAC1. These studies suggest that EAAC1 and GTRAP3-18 can interact *in vivo*.

GTRAP3-18 messenger RNA was widely expressed in brain regions and body organs, consistent with the distribution of EAAC1 (Fig. 2a)^{8,9}. Similarly, GTRAP3-18 protein was expressed in many neural and non-neural tissues, when examined using a polyclonal oligopeptide antibody to the N terminus of GTRAP3-18 (Fig. 2b). GTRAP3-18 protein seemed to aggregate as multimers.

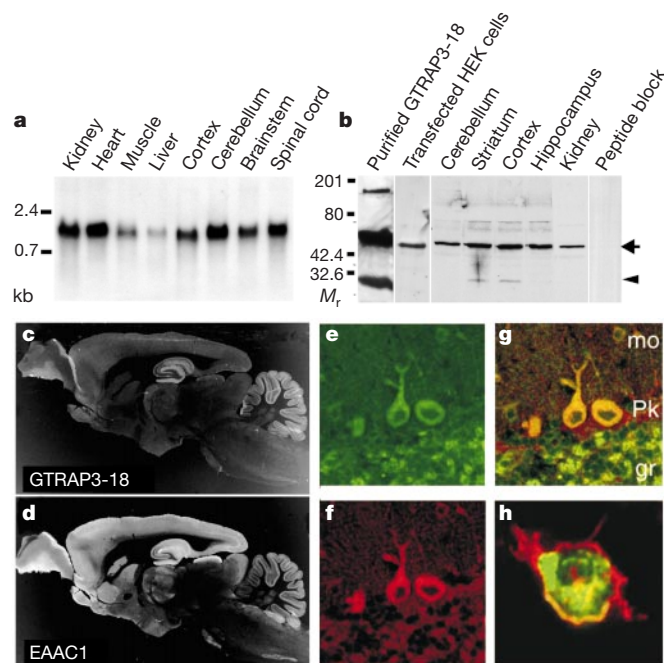


Figure 2 Tissue and cellular distribution of GTRAP3-18 protein and mRNA. **a**, Northern blot analysis of GTRAP3-18 mRNA. **b**, Western blot analyses of GTRAP3-18 protein. Pure GTRAP3-18 (~30K, arrowhead) monomer tended to form multimers including the dimer (arrow). Dimeric GTRAP3-18 was found in HEK293 cells transfected with GTRAP3-18 cDNA and tissue homogenates. **c**, **d**, Immunostaining of GTRAP3-18 and EAAC1 proteins in rat brain. **e**–**h**, Confocal microscopy of cellular colocalization of GTRAP3-18 and EAAC1. In the cerebellum, GTRAP3-18 (**e**, green), EAAC1 (**f**, red) reveals prominent cytosolic colocalization (**g**, yellow). mo, molecular layer; pk, Purkinje cell layer; gr, granular layer. In transfected HEK293 cells (**h**), EAAC1 (red) tends to localize to the cell surface whereas GTRAP3-18 was found typically cytosolic and co-localized with EAAC1 (yellow) at the cell membrane.

The dimeric form of GTRAP3-18 was the predominant species in tissue homogenates, and it was also observed when purified GTRAP3-18 protein was detected using the N-terminal GTRAP3-18 antibody (Fig. 2b, lane 1), and when c-Myc-GTRAP3-18 protein was detected using anti-c-Myc antibodies (Fig. 2b, lane 2).

Immunohistological analysis of rat brain revealed that GTRAP3-18 protein was expressed widely (Fig. 2c) and primarily localized to neurons such as cerebellar Purkinje cells (Fig. 2e, g), identical to the expression and localization⁴ of EAAC1 (Fig. 2f, g)². In transfected HEK293 cells, EAAC1 protein (red) seemed to be aggregated at the cell membrane (Fig. 2h), whereas GTRAP3-18 protein (green) was typically localized to the cell membrane and cytosol, and co-associated with EAAC1 protein (yellow) at the cell membrane.

We tested whether GTRAP3-18 modulates EAAC1 function by studying, 72 h after transfection, sodium-dependent [³H]glutamate transport^{3,10} in HEK293 cells co-expressing both proteins. Total glutamate transport decreased progressively with increasing expression of GTRAP3-18 protein (Fig. 3a). This effect was specific for EAAC1; co-expression of GTRAP3-18 with EAAT4 had no effect on transport activity. The inhibition of transport was not caused by a decrease of EAAC1 protein level by the co-expression of GTRAP3-18, as quantitated by western blot (Fig. 3b). Similarly, the loss of EAAC1 activity was not due to altered protein trafficking. Even at high levels of GTRAP3-18 expression—when little EAAC1-mediated transport was observed—surface biotinylation and confocal microscopy showed that EAAC1 surface expression was unaltered (Fig. 3b).

To evaluate the biochemical nature of altered transport, we carried out kinetic analyses with HEK293 cells co-expressing EAAC1 and GTRAP3-18. EAAC1 and GTRAP3-18 co-expressing cells showed a decrease in affinity (Michaelis constant, $K_m = 40 \mu\text{M}$, $n = 4$, $P < 0.01$) without a shift in maximal velocity ($V_{\max} = 0.99 \text{ nmol min}^{-1} \text{ per mg protein}$), when compared with cells only expressing EAAC1 ($K_m = 9 \mu\text{M}$; $V_{\max} = 1.02 \text{ nmol min}^{-1} \text{ per mg protein}$; Fig. 3c). Similar results were observed with other cell lines (COS7 and C6 glioma; data not shown).

These studies indicate that GTRAP3-18 modulates EAAC1 transport activity by decreasing its affinity for glutamate. On the basis of these results, we thought that GTRAP3-18 might tonically modulate EAAC1 activity. To test this, we used antisense oligomers to lower GTRAP3-18 expression in HEK293 cells. Notably, western blot analyses and glutamate uptake assays revealed that HEK293 cells endogenously express some EAAC1 and GTRAP3-18 protein, but do not express the transporter subtypes GLAST, GLT-1 or EAAT4. We transfected antisense oligomers, targeted to the 5' GTRAP3-18 transcript, into HEK293 cells. Antisense oligomers specifically reduced endogenous GTRAP3-18 protein level (Fig. 3d, grey bars); EAAC1 protein level was not affected. Significantly, glutamate transport activity was elevated concomitantly with the reduction of GTRAP3-18 protein level (black bars).

We determined whether the *in vitro* modulation of EAAC1 by GTRAP3-18 is physiologically relevant, by administering GTRAP3-18 antisense oligomers intraventricularly. Eleven days of antisense treatment resulted in a reduction of GTRAP3-18 protein level and a significant increase in cortical glutamate uptake, whereas glutamate uptake was not altered in animals treated with the sense oligomer (Fig. 3e). The effect was due to increased EAAC1-mediated transport because it was not altered by dihydrokainic acid (DHK), an inhibitor of GLT-1-mediated glutamate transport¹¹. In kinetic studies of DHK-insensitive, cortical glutamate uptake from antisense-treated animals, the apparent affinity for glutamate was increased (antisense $K_m = 10 \mu\text{M}$; $V_{\max} = 1.08 \text{ nmol min}^{-1} \text{ per mg protein}$) compared with artificial cerebrospinal-fluid (CSF)-treated or sense-treated control animals (control $K_m = 19.7 \mu\text{M}$; $V_{\max} = 1.08 \text{ nmol min}^{-1} \text{ per mg protein}$; Fig. 3f). These results suggest that GTRAP3-18 negatively modulates EAAC1 glutamate transport activity *in vivo*.

Human GTRAP3-18 (JWA protein), as described above, was originally identified as a retinoic-acid-responsive gene. We therefore tested whether retinoic acid could upregulate GTRAP3-18 expression and consequently inhibit EAAC1-mediated glutamate transport in HEK293 cells. Retinoic acid induced a large increase in GTRAP3-18 expression, over a non-toxic dose range from 1 to 10 μ M (Fig. 4a). A significant decrease in glutamate transport activity paralleled the increase of GTRAP3-18 protein level (Fig. 4a). Fluorescence microscopy indicated that this loss of transport activity was not due to changes in EAAC1 protein level (Fig. 4a) or in the cellular membrane localization of EAAC1 protein induced by retinoic acid (Fig. 4b).

To confirm that loss of transport activity was specifically due to GTRAP3-18 and not by other factors induced by retinoic acid or direct effects on EAAC1, we constructed a truncated EAAC1 cDNA,

lacking the last 93 amino acids. The truncation corresponded to the region used as bait in yeast two-hybrid screening, and was not able to interact with GTRAP3-18 (Fig. 1b, lane 5). Nevertheless, after transient expression in HEK293 cells, the truncated EAAC1 transported glutamate. Retinoic acid treatment did not alter activity of the truncated EAAC1 protein; even though GTRAP3-18 protein expression was elevated markedly (Fig. 4c). Thus, the loss of transport activity caused by retinoic acid was the result of GTRAP3-18 induction. Notably, truncated EAAC1 had increased glutamate transport activity as compared with wild type. Truncated EAAC1 had a K_m of 5.4 μ M, which was more than a threefold increase in affinity over that of wild-type EAAC1 ($K_m = 17 \mu$ M; Fig. 4d). This might mean that endogenous EAAC1 is normally inhibited by GTRAP3-18—an effect that was eliminated in a truncated EAAC1, and that was mimicked by GTRAP3-18 antisense

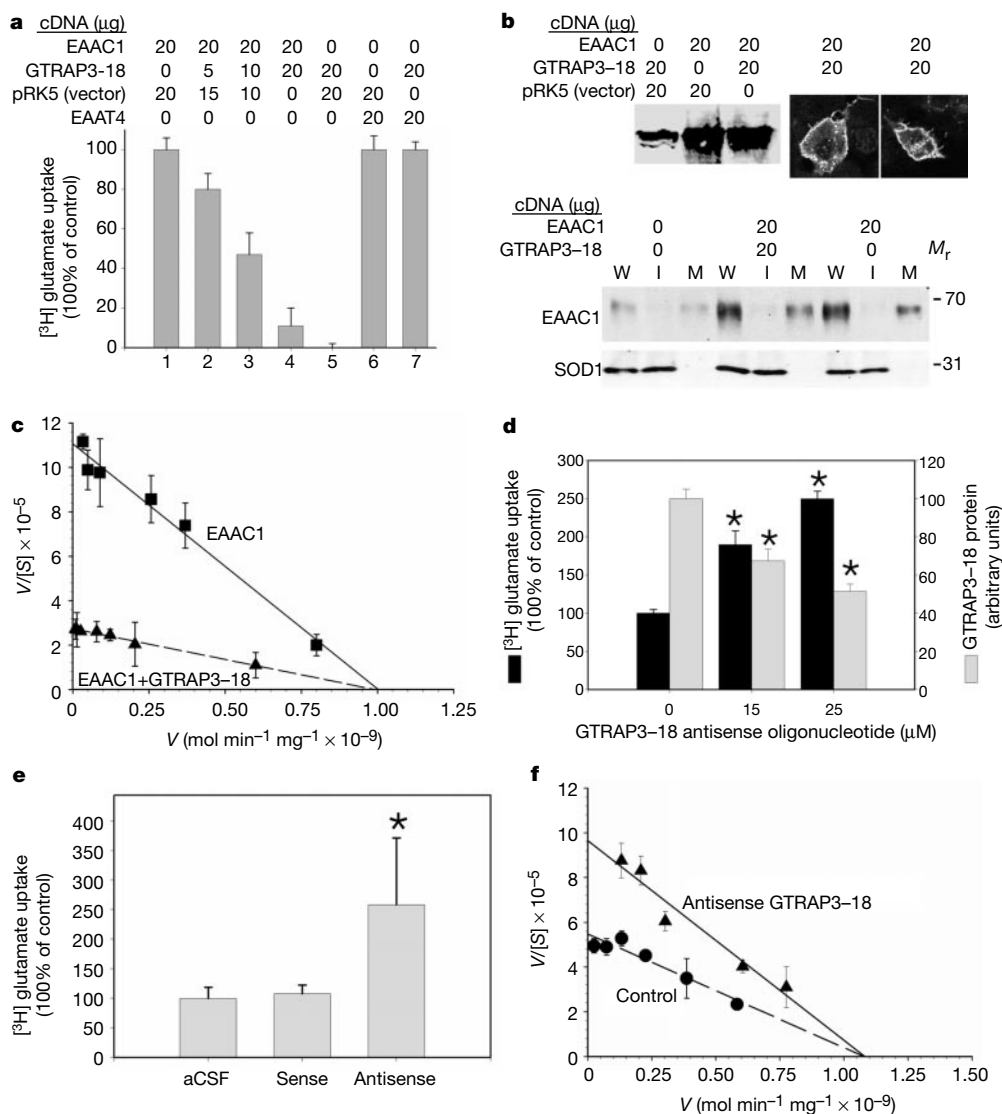


Figure 3 GTRAP3-18 negatively modulates EAAC1-mediated glutamate transport. **a**, Glutamate transport in HEK293 cells transfected with the indicated plasmids. GTRAP3-18 inhibited EAAC1-mediated transport, but had no effect on EAAT4 ($n = 6$). **b**, Western blot showed that the co-expression of GTRAP3-18 had no effect on total EAAC1 protein expression (top, left). Analysis of HEK293 cells by confocal microscopy (top panel, right) and surface biotinylation (bottom) revealed no alteration in the membranous localization of EAAC1. I, intracellular fraction; M, cellular membrane fraction; W, whole cell. Superoxide dismutase (SOD1) was used as a control. **c**, Edie-Scatchard plot of glutamate transport in transfected HEK293 cells showed a 4–10-fold decrease in affinity ($n = 4$). **d**, Antisense

GTRAP3-18 oligomers reduced endogenous GTRAP3-18 expression in HEK293 cells and produced an increase in EAAC1 transport. Asterisks indicate statistical significance (Students t -test; $P < 0.01$) versus untreated control ($n = 4$). **e**, Intraventricular antisense delivery led to a 50% loss of cortical GTRAP3-18 levels and a significant increase in dihydrokainate-insensitive glutamate uptake. Asterisk indicates statistical significance (Students t -test; $P < 0.05$) versus artificial CSF (aCSF) control ($n = 7$). **f**, Kinetic analysis of cortical tissue from rats treated with antisense GTRAP3-18 shows an increase in apparent affinity for glutamate. All errors bars represent standard deviation.

treatment (Fig. 3f).

To test this hypothesis *in vivo*, retinoic acid was infused intravenously. After 4 d of treatment, cortical GTRAP3-18 protein expression was increased in a dose-dependent manner, and this was associated with a significant decrease of total glutamate uptake (Fig. 4e, top panel). This effect was specifically due to decreased EAAC1-mediated transport because it was not altered by the glutamate transport inhibitor dihydrokainic acid, at a concentra-

tion that predominantly effects GLT-1 (Fig. 4e, bottom panel)¹¹. Kinetic analysis of DHK-insensitive, cortical glutamate transport from animals treated 4 d with intraventricular retinoic acid revealed a fourfold decrease in affinity over control transport (Fig. 4f)—very similar to that seen *in vitro* (Fig. 3c). In addition, retinoic acid inhibition of glutamate transport could be reversed *in vivo*; chronic intraventricular treatment with antisense GTRAP3-18 oligomer (50–100 ng d⁻¹, for 7–10 d) blocked the retinoic acid (2.5 μ M)

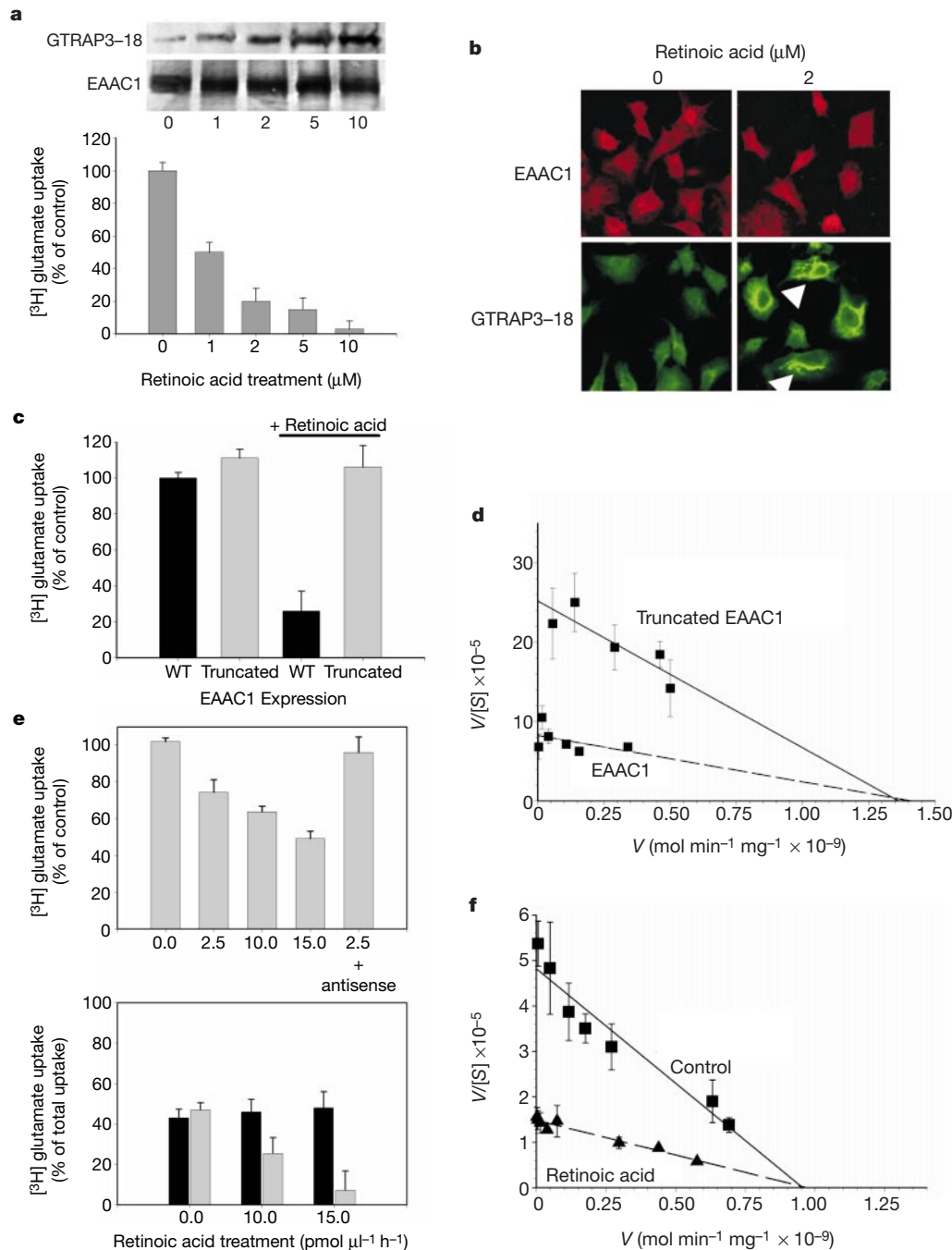


Figure 4 Retinoic acid upregulates GTRAP3-18 expression and consequently inhibits EAAC1 transport in HEK293 cells in brain. **a**, After 14 d, retinoic acid induced GTRAP3-18 expression and inhibited transport ($n = 5$) in HEK293 cells. **b**, Membrane localization of EAAC1 protein was not changed by retinoic acid, as visualized by fluorescent microscopy. GTRAP3-18 expression, particularly in a subcellular compartment, was greatly induced by retinoic acid (arrow). **c**, Glutamate transport by HEK293 cells expressing truncated EAAC1 (lacking a C-terminal interacting domain) was not inhibited by 14 d of treatment with retinoic acid (5 μ M; $n = 3$). **d**, Truncated EAAC1 expressed by HEK293 cells has

increased apparent affinity for glutamate compared with wild type (WT) EAAC1. **e**, Intraventricular delivery of retinoic acid (3–5 d) inhibited glutamate uptake; the effect was blocked by concomitant treatment with GTRAP3-18 antisense oligomers (5 μ g μ l⁻¹; 1 μ l h⁻¹; top panel). This effect was specific for DHK-insensitive transport (bottom panel, grey bars)—a crude partial estimate of EAAC1 activity *in vivo*—but not for DHK-sensitive transport (EAAT2-mediated) (black bars; $n = 5$). **f**, Kinetic analysis of cortical uptake from rats treated with retinoic acid (2.5 pmol l⁻¹ h⁻¹ for 4 d) showed a large decrease in apparent affinity for glutamate ($n = 3$). All errors bars represent standard deviation.

induction of GTRAP3-18, and also blocked the inhibition of glutamate transport seen with retinoic acid treatment (Fig. 4e, top panel). Retinoic acid had no effect on glutamate transport by cells expressing GLT-1 or EAAT4.

In brain, glutamate serves as an excitatory neurotransmitter, a metabolic substrate for other neurotransmitters (that is, GABA (γ -amino butyric acid)), and an amino acid for general cellular metabolism. In brain, glutamate transporters maintain low extracellular glutamate and influence the kinetics of glutamate receptor activation^{12–14}. Little is known about the molecular and protein regulation of glutamate transporters, or for neurotransmitter transporters in general. In tumour cell lines, for example, the cell-surface expression of EAAC1 appears to be regulated by pathways mediated by both protein kinase C and phosphatidylinositol-3-OH kinase, albeit through as yet unidentified proteins^{7,15}. GTRAP3-18 appears to be an endogenous inhibitory regulator of EAAC1. EAAC1 is found throughout the brain on somas and dendrites of small and large pyramidal neurons^{2,4}. EAAC1 is also localized to pre-synaptic GABA-containing terminals, and may have a metabolic role in providing glutamate for GABA metabolism^{4,16}. Loss of brain EAAC1 expression interferes with GABA synthesis and results in epilepsy^{3,17}. In kidney, EAAC1 may contribute to renal acidic amino-acid reabsorption, acid/base balance, cell volume regulation, and amino-acid metabolism^{5,6}. GTRAP3-18 may have a critical role in the regulation of the neurotransmitter and metabolic functions of EAAC1. □

Methods

Yeast two-hybrid screening

The MATCHMAKER Two-Hybrid System (Clontech) was used for screening. We screened a rat brain cDNA library with a bait protein corresponding to C-terminal intracellular domain of EAAC1 (last 87 amino acids, cDNA position 1,458–1,719). The positive clones were selected and assayed for β -galactosidase activity. Plasmid DNAs were isolated from positive clones and co-transformed with bait cDNA back into yeast to reconfirm the interaction.

GST fusion proteins

The GST Gene Fusion System (Pharmacia) was used to construct and generate GST–EAAC1 and GST–GTRAP3-18 fusion proteins using pGEX–6P-1 vector.

In vitro cell-free binding assay

The TnT Coupled Reticulocyte Lysate System (Promega) was used to generate [³⁵S]methionine-labelled GTRAP3-18 protein using the pRK5–GTRAP3-18 plasmid DNA as a template, driven by SP6 RNA polymerase. GST–EAAC1 fusion proteins were expressed in *Escherichia coli* and recovered on glutathione–sepharose beads as described above. The beads were incubated with lysate which containing [³⁵S]labelled GTRAP3-18 protein for 1 h at 4 °C and then washed five times with NETN buffer (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40). The beads were boiled in sample buffer (SB) (2% SDS, 10% glycerol, 62 mM Tris pH 6.8) and bound proteins resolved in SDS–PAGE.

Gene constructions and mammalian cell expression system

We used the eukaryotic expression vectors pcDNA3 and pRK5 for expression of cDNAs in mammalian cell lines. Full-length EAAC1 cDNA was subcloned into *NotI*, *EcoRI* sites of the pcDNA3 vector. For the truncated EAAC1, the last 279 nucleotides were removed by *AatII* and *EcoRI* digestion in a pcDNA3/EAAC1 construct and blunt ended. For tagging GTRAP3-18 with c-Myc, GTRAP3-18 cDNA was first subcloned into downstream of c-myc gene (*NotI* site) in the pTYGL vector. The *EcoRI*/*BamHI* fragment containing Myc–GTRAP3-18 fusion gene was then subcloned into pRK5. The constructed plasmid DNA was transfected into HEK293, COS-7 or C6 glioma cell lines by standard electroporation, and 72 h later cells were collected for various assays.

Immunoprecipitation, biotinylation and transport assays

For the transfected cells, cells were solubilized in 1 ml of IPB buffer (10 mM Tris pH 7.6, 5 mM EDTA, 5 mM EGTA, 1 mM sodium orthovanadate, aprotinin 15 μ g ml^{–1}, 0.1 mM PMSF) with 1% triton at 4 °C for 30 min with gentle rotation. Cell debris was removed (15,000g, 15 min), and supernatants were incubated (4 °C, 16 h) with protein A/sepharose beads (150 μ l, CL-4B, Amersham) and primary antibodies (1.5 mg ml^{–1}). The immunobead-bound protein complexes were washed three times (IPB buffer in 1% triton X-100, 100 mM NaCl) followed by three IPB buffer washes. Immunoprecipitated proteins were boiling in SB and analysed by western blotting.

For the rat brain, coronal sections of brain were sliced at 1–2-mm intervals from the cerebellum to the olfactory bulbs. The cortex region was excised from the brain and placed in cold buffer A (50 mM Tris pH 7.5, 2 mM EDTA, 150 mM NaCl, 0.5 mM dithiothreitol).

The tissue was washed three times in buffer A and then weighed. The tissue was then homogenized using a blender in 2.5 volumes of buffer B (50 mM Tris pH 7.5, 10% glycerol, 5 mM magnesium acetate, 0.2 mM EDTA, 0.5 mM dithiothreitol, 1 mM PMSF). The particulate material was removed by centrifugation at 15,000g for 30 min at 4 °C. The supernatant fraction was incubated with protein A/sepharose beads and primary antibodies as above.

Biotinylation was performed as described¹⁵ with modifications. Glutamate uptake assays were done as described on either cells grown on six-well plates (up to 1 mg protein) or brain tissues (100–150 μ g protein). To block GLT-1 transport activity, homogenates were pre-incubated with dihydrokainic acid (300 μ M, Sigma) for 20–60 min before assay.

Blotting, immunostaining and antibodies

The methods for isolation of RNA and northern blotting, western blotting and immunostaining have been described¹⁸. We synthesized a synthetic peptide corresponding to an N-terminal domain of GTRAP3-18 protein NH₂-K¹⁴FFPGSDRFARPD²⁸-COOH and used this to generate polyclonal, affinity-purified antibodies from New Zealand White rabbits^{2,19}. Confocal microscopy of transfected cells of brain sections was performed with a Zeiss LSM 510 laser scanning microscope using fluorescein (Vector, FI1000) and Texas red (Vector, TI2000) fluorochromes.

Intraventricular antisense/drug administration

We used the following sequences for the new phosphodiester oligonucleotides: sense GTRAP3-18, 5'-GTGAACCTTGCCCGCTC-3'; antisense GTRAP3-18, 5'-GAGCGG GCGAAGGTTACAC-3'. Oligonucleotides (5 μ g μ l^{–1}), retinoic acid (1–20 μ M; 0–20 pmol μ l^{–1}) separately or in combination were administered intraventricularly over 3–11 d, by mini-osmotic pumps (Alza Corporation, Palo Alto, CA) as described³.

Received 5 October; accepted 20 November 2000.

1. Kanai, Y. & Hediger, M. A. Primary structure and functional characterization of a high-affinity glutamate transporter. *Nature* **360**, 467–471 (1992).
2. Rothstein, J. D. *et al.* Localization of neuronal and glial glutamate transporters. *Neuron* **13**, 713–725 (1994).
3. Rothstein, J. D. *et al.* Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate. *Neuron* **16**, 675–686 (1996).
4. He, Y., Janssen, W. G. M., Rothstein, J. D. & Morrison, J. H. Differential synaptic localization of the glutamate transporter EAAC1 and glutamate receptor subunit GluR2 in the rat hippocampus. *J. Comp. Neurol.* **418**, 255–269 (2000).
5. Shaykul, C. *et al.* Localization of the high-affinity glutamate transporter EAAC1 in rat kidney. *Am. J. Physiol. Renal Physiol.* **42**, F1023–F1029 (1997).
6. Peghini, P., Janzen, J. & Stoffel, W. Glutamate transporter EAAC1-deficient mice develop dicarboxylic aminoaciduria and behavioral abnormalities but no neurodegeneration. *EMBO J.* **16**, 3822–3832 (1997).
7. Sims, K. D., Straff, D. J. & Robinson, M. B. Platelet-derived growth factor rapidly increases activity and cell surface expression of the EAAC1 subtype of glutamate transporter through activation of phosphatidylinositol 3-kinase. *J. Biol. Chem.* **275**, 5228–5237 (2000).
8. Hediger, M. A. & Welbourne, T. C. Introduction: glutamate transport, metabolism, and physiological responses. *Am. J. Physiol.* **277**, F477–F480 (1999).
9. Hediger, M. A. Glutamate transporters in kidney and brain. *Am. J. Physiol.* **277**, F487–F492 (1999).
10. Lin, C. G. *et al.* Aberrant RNA processing in a neurodegenerative disease: The cause for absent EAAT2 a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron* **20**, 589–602 (1998).
11. Robinson, M. B. The family of sodium-dependent glutamate transporters: a focus on the GLT-1/EAAT2 subtype. *Neurochem. Int.* **33**, 479–491 (1998).
12. Bergles, D. E. & Jahr, C. E. Synaptic activation of glutamate transporters in hippocampal astrocytes. *Neuron* **19**, 1297–1308 (1997).
13. Bergles, D. E., Dzubay, J. A. & Jahr, C. E. Glutamate transporter currents in bergmann glial cells follow the time course of extrasynaptic glutamate. *Proc. Natl Acad. Sci. USA* **94**, 14821–14825 (1997).
14. Diamond, J. S. & Jahr, C. E. Transporters buffer synaptically released glutamate on a submillisecond time scale. *J. Neurosci.* **17**, 4672–4687 (1997).
15. Davis, K. E. *et al.* Multiple signaling pathways regulate cell surface expression and activity of the excitatory amino acid carrier 1 subtype of Glu transporter in C6 glioma. *J. Neurosci.* **18**, 2475–2485 (1998).
16. Conti, F., DeBiasi, S., Minelli, A., Rothstein, J. D. & Melone, M. EAAC1, a high-affinity glutamate transporter, is localized to astrocytes and GABAergic neurons besides pyramidal cells in the rat cerebral cortex. *Cereb. Cortex* **8**, 108–116 (1998).
17. Sepkuty, J., Eccles, C. U., Lesser, R. P., Dykes-Hoberg, M. & Rothstein, J. D. Molecular knockdown of neuronal glutamate transporter EAAT3 produces epilepsy and dysregulation of GABA metabolism. *Soc. Neurosci. Abstracts* **23**, 1484 (1997).
18. Lin, C. L. G. *et al.* Molecular cloning and expression of the rat EAAT4 glutamate transporter subtype. *Mol. Brain Res.* **63**, 174–179 (1998).
19. Furuta, A., Martin, L. J., Lin, C. L. G., Dykes-Hoberg, M. & Rothstein, J. D. Cellular and synaptic localization of the neuronal glutamate transporters excitatory amino acid transporter 3 and 4. *Neuroscience* **81**, 1031–1042 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Huganir for the pRK5 vector; J. Sepkuty, R. Ganel, and W. Song for helpful suggestions and discussions; and L. Jin, C. Coccia and B. Kim for technical support.

Correspondence and requests for materials should be addressed to J.D.R. (e-mail: jrothste@welchlink.welch.jhu.edu). GenBank accession number for GTRAP3-18 is AF240182.

Modulation of the neuronal glutamate transporter EAAT4 by two interacting proteins

Mandy Jackson*, Wei Song*, Mu-Ya Liu†, Lin Jin*, Margaret Dykes-Hoberg‡, Chien-liang G. Lin*, William J. Bowers‡, Howard J. Federoff‡, Paul C. Sternweis† & Jeffrey D. Rothstein*

* Department of Neurology and Neuroscience, Johns Hopkins University, Baltimore, Maryland 21287, USA

† Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

‡ Department of Neurology, and the Center for Aging and Developmental Biology, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, USA

Glutamate is the main excitatory neurotransmitter in the mammalian central nervous system and is removed from the synaptic cleft by sodium-dependent glutamate transporters. To date, five distinct glutamate transporters have been cloned from animal and human tissue: GLAST (EAAT1), GLT-1 (EAAT2), EAAC1 (EAAT3), EAAT4, and EAAT5 (refs 1–5). GLAST and GLT-1 are localized primarily in astrocytes^{6,7}, whereas EAAC1 (refs 8, 9), EAAT4 (refs 9–11) and EAAT5 (ref. 5) are neuronal. Studies of EAAT4 and EAAC1 indicate an extrasynaptic localization on perisynaptic membranes that are near release sites^{8–10}. This localization facilitates rapid glutamate binding, and may have a role in shaping the amplitude of postsynaptic responses in densely packed cerebellar

terminals^{12–15}. We have used a yeast two-hybrid screen to identify interacting proteins that may be involved in regulating EAAT4—the glutamate transporter expressed predominately in the cerebellum—or in targeting and/or anchoring or clustering the transporter to the target site. Here we report the identification and characterization of two proteins, GTRAP41 and GTRAP48 (for glutamate transporter EAAT4 associated protein) that specifically interact with the intracellular carboxy-terminal domain of EAAT4 and modulate its glutamate transport activity.

To identify proteins that interact with the C terminus of the EAAT4 protein, we used the last 77 amino acids of EAAT4 as bait to screen a rat brain complementary DNA library. We isolated two independent cDNA clones and called them GTRAP41 and GTRAP48 (for glutamate transporter-4-associated protein). Isolation of the full-length cDNAs by a series of 5' and 3' rapid amplification of cDNA ends (RACE) polymerase chain reactions (PCRs) showed that the largest open reading frame (ORF) for GTRAP41 is 7,164 base pairs (bp), which encodes a 2,388 amino-acid protein with a predicted relative molecular mass (M_r) of 270,958 (accession no. AF225960).

A BLAST search of the GenBank database showed that GTRAP41 possesses 87% identity with β -spectrin III (accession no. AB008567). GTRAP41 possesses seventeen 16-amino-acid spectrin repeats, two α -actinin domains and a pleckstrin homology (PH) domain (Fig. 1a). The largest ORF identified for GTRAP48 (accession no. AF225961) is 4,581 bp, which encodes a 1,527-amino-acid protein with a predicted M_r of 168,698. A BLAST search of the GenBank database showed that GTRAP48 is unique, but it possesses significant homology to the KIAA0380 cDNA encoded protein (90% identity) and RhoGEF, p115 (ref. 16). GTRAP48 possesses a PDZ domain, a regulatory G-protein-signalling sequence (LH),

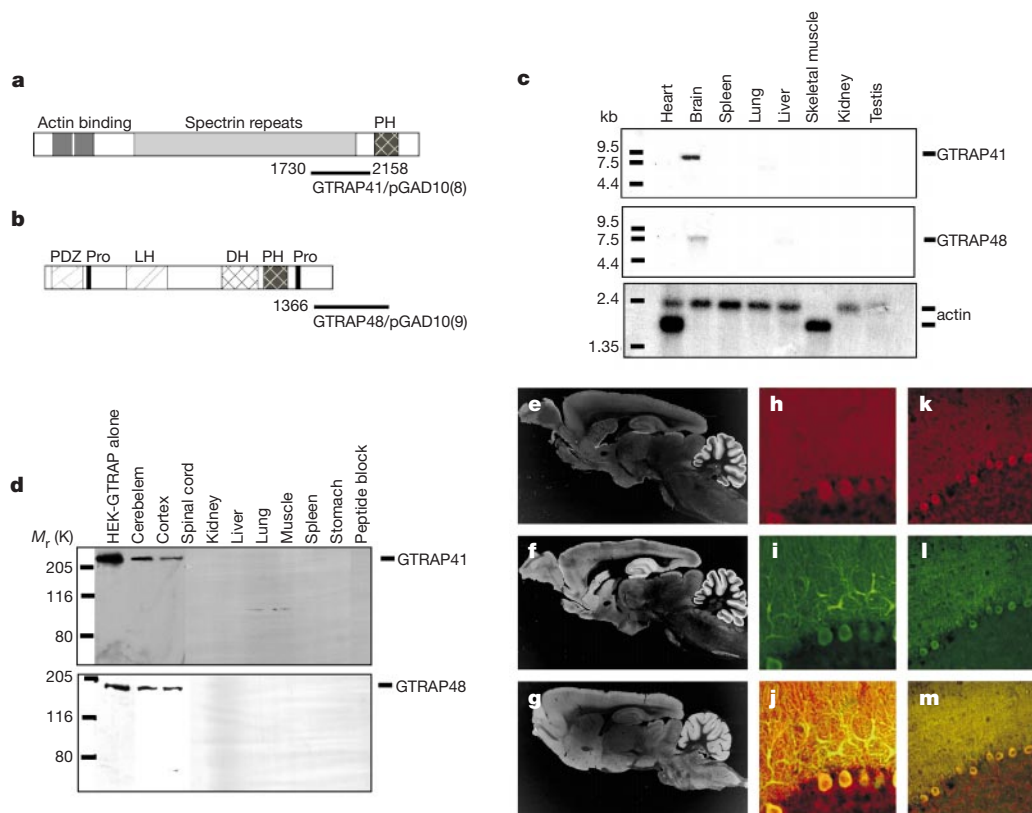


Figure 1 Structure and distribution of GTRAP41 and GTRAP48. **a, b**, cDNA clones GTRAP41/pGAD10 and GTRAP48/pGAD10 isolated from the yeast two-hybrid screen are shown aligned below representations of full-length GTRAP41 and GTRAP48, respectively. The number of times the clones were isolated is shown in parentheses. **c**, Multiple tissue northern (MTN; Clontech) blot probed with 3' PCR probes of GTRAP41 and GTRAP48.

d, Western analysis of GTRAP41 and GTRAP48. **e–g**, Rat brain sections stained with anti-EAAT4 (**e**), anti-GTRAP41 (**f**) and anti-GTRAP48 (**g**) antibodies. **h–m**, EAAT4 (**h, k**), GTRAP41 (**i**) and GTRAP48 (**l**) are all predominately expressed in the cell bodies and dendrites of Purkinje cells. The overlaps of GTRAP41 and GTRAP48 with EAAT4 are shown in **j** and **m**, respectively.

tandem dbl homology (DH) and pleckstrin homology (PH) domains characteristic of guanine nucleotide exchange factors for the Rho family of G proteins, and two proline-rich sequences (Fig. 1b).

Northern blot analysis detected a 8.3-kilobase (kb) GTRAP41 and a 7.5-kb GTRAP48 messenger RNA in brain tissue (Fig. 1c). Longer exposure revealed a low level of expression in liver and kidney for both GTRAP41 and GTRAP48. Anti-peptide antibodies were raised and the affinity-purified antibodies recognized a M_r 270,000 (270K) protein and a 170K protein, respectively, in HEK 293T cells transfected with full-length GTRAP41 and GTRAP48 cDNA (Fig. 1d). GTRAP41 and GTRAP48 were both selectively localized to brain, consistent with the northern blot analysis (Fig. 1d). EAAT4 is selectively localized to cerebellar Purkinje cells, although low-level expression is observed in cerebral cortex, hippocampus and striatum⁹. We found that GTRAP41 and GTRAP48 were expressed predominately in the cerebellum (Fig. 1e–g), with low-level immunoreactivity in striatum, hippocampus and thalamus. Immunofluorescence microscopy revealed that all three proteins are expressed in cerebellar Purkinje cell soma and dendrites, with little axonal staining (Fig. 1h–m).

We first confirmed the biochemical interaction between GTRAP41/GTRAP48 and EAAT4 by an *in vitro* binding assay. Full-length Myc-tagged GTRAP41 and GTRAP48 were expressed in HEK 293T cells. The solubilized cell extracts were then mixed with bead-linked glutathione *S*-transferase (GST)–EAAT4 or GST alone, and the bound proteins were eluted. GTRAP41 and GTRAP48 were retained specifically by the GST–EAAT4 fusion protein, but not by GST alone (Fig. 2a). A stable, rat EAAT4-expressing cell line was generated in HEK 293T cells (HEK-rEAAT4) and transfected with cDNAs encoding Myc-tagged GTRAP41 and GTRAP48. We used antibodies directed at the amino terminus of EAAT4 to immunoprecipitate the antigen plus any associated protein. Western blot analysis using an anti-c-Myc antibody showed that GTRAP41 and GTRAP48 co-immunoprecipitated with EAAT4 (Fig. 2b).

Similarly, when the anti-c-Myc antibody was used, EAAT4 was co-immunoprecipitated with GTRAP41 and GTRAP48 (Fig. 2c). *In vivo*, GTRAP41 (Fig. 2d) and GTRAP48 (Fig. 2f) were co-immunoprecipitated with EAAT4 from brain by antibodies directed at the N terminus of EAAT4 but not by antibodies directed at the C terminus (Fig. 2d, f). As the site of interaction is within the C terminus of EAAT4, however, it is likely that the C-terminal antibodies disrupt the protein–protein interaction. Furthermore, GTRAP41 and GTRAP48 seem to interact specifically with EAAT4, as we could not co-immunoprecipitate GTRAP41 and GTRAP48 from brain with antibodies directed at other glutamate transporters (Fig. 2d, f).

Western blot analysis confirmed that the immunoprecipitating antibodies pulled down their corresponding antigen (see Supplementary Information). No co-immunoprecipitation was observed when the precipitating antibody was omitted or pre-absorbed (Fig. 2b, e, g). We next determined the region of EAAT4 that binds GTRAP41 and GTRAP48. We used a series of successively larger C-terminal deletions of the original, 77-amino-acid, C-terminal EAAT4 bait in a yeast two-hybrid screen. Residues 555–561 and 527–534 appear to be required for GTRAP41 and GTRAP48 binding, respectively.

As GTRAP48 possesses areas of homology to p115 and PDZ-RhoGEF, two RhoGEFs that selectively activate Rho^{16,17}, we investigated whether GTRAP48 interacts with the Rho family of GTPases. We measured the amount of GTP- γ S that bound to GST–RhoA, GST–Cdc42 and GST–Rac in the presence of full-length GTRAP48 or p115, and found that GTRAP48, like p115, shows a specific guanine nucleotide exchange activity for Rho (Fig. 3a, b). Co-immunoprecipitation assays also showed that GTRAP48 interacts with the active form (in the presence of aluminium fluoride) but not the inactive form of the $G\alpha_{13}$ subunit and therefore, may act as a link between G-protein-coupled receptors and their downstream targets (Fig. 3c). Specificity of the $G\alpha_{13}$ antibody has been previously described¹⁸.

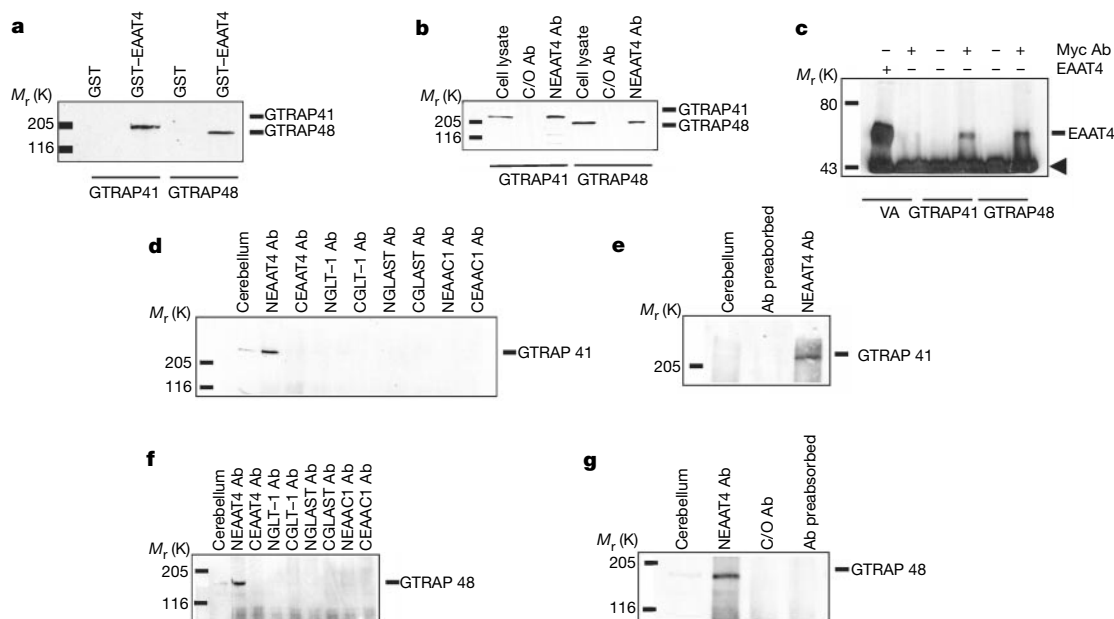


Figure 2 Interaction of GTRAP41 and GTRAP48 with EAAT4. **a**, Binding of Myc–GTRAP41 and Myc–GTRAP48 to GST–EAAT4. **b**, **c**, GTRAP41, GTRAP48 or vector alone (VA) were expressed (as indicated by the bars) in HEK-rEAAT4 cells. Immunoprecipitations were performed either with the N-terminal anti-EAAT4 antibody (**b**), or with the antibodies (Ab) indicated above the lanes (**c**). Immunoprecipitates were analysed by immunoblotting using anti-c-Myc (**b**) or C-terminal anti-EAAT4 (**c**; arrowhead indicates protein-A–

Sepharose band) antibodies. **d–g**, Extracts of rat brain were immunoprecipitated with antibodies (**d**, **f**) directed at the N terminus and C terminus of the glutamate transporters (as indicated above the lanes), no antibody or antibody pre-absorbed with peptide (**e**, **g**). Immunoblots were probed with the anti-GTRAP41 (**d**, **e**) and the anti-GTRAP48 (**f**, **g**) antibodies.

Rho regulates the remodelling of the actin cytoskeleton through various actin-binding proteins, although the mechanism is not well characterized¹⁹. As GTRAP48 can activate Rho, we next determined whether the expression of GTRAP48 induced reorganization of the actin cytoskeleton and whether it altered the distribution of GTRAP41, a possible actin-binding protein. When GTRAP41 was expressed alone there was a close relationship between actin and GTRAP41 at the cell membrane, but there were very few organized

actin filaments (Fig. 3d, e). Conversely, when GTRAP41 and GTRAP48 were co-expressed, GTRAP41 colocalized with actin in structures that resembled actin-stress fibres (Fig. 3f, g), a typical Rho-dependent effect. We noted that the overexpression of GTRAP48 also induced the formation of membrane ruffling and filopodia (Fig. 3h, i), suggesting some degree of cross-talk between the small GTPases, as these are typical Rac- and Cdc42-dependent effects.

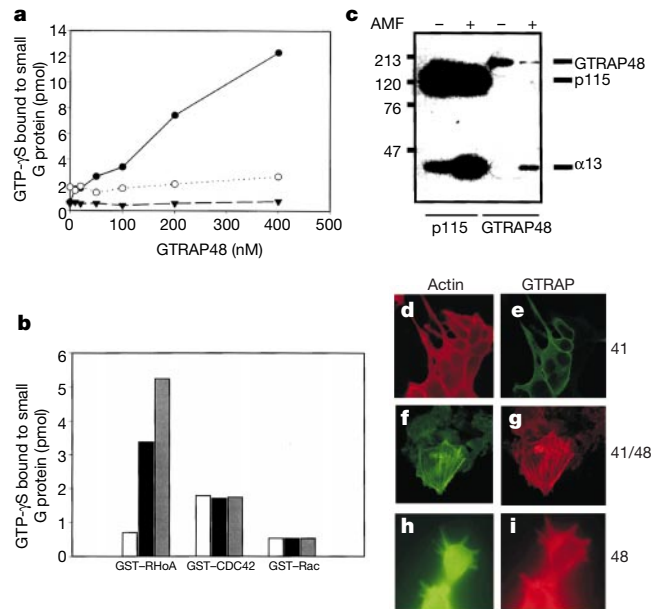


Figure 3 Guanine nucleotide exchange activity of GTRAP48. **a**, Binding of GTP- γ S to GST-RhoA (filled circles), GST-CDC42 (open circles) and GST-Rac (filled triangles) was measured after 10 min at 30 °C in the presence of the indicated concentrations of full-length GTRAP48 as described in Methods. **b**, Binding of GTP- γ S to the indicated GTPases in either the absence (white), or the presence of 100 nM GTRAP48 (black) or 100 nM

p115-RhoGEF (grey). **c**, Binding of active $G\alpha_{13}$ to Glu-tagged (EE) GTRAP48 and p115 RhoGEF. **d–i**, HEK-rEAAT4 cells transfected with either vector alone (**d**, **e**), Myc-tagged GTRAP41 and GTRAP48 (**f**, **g**) or Myc-tagged GTRAP48 (**h**, **i**). Actin filaments were visualized with FITC- or rhodamine-conjugated phalloidin.

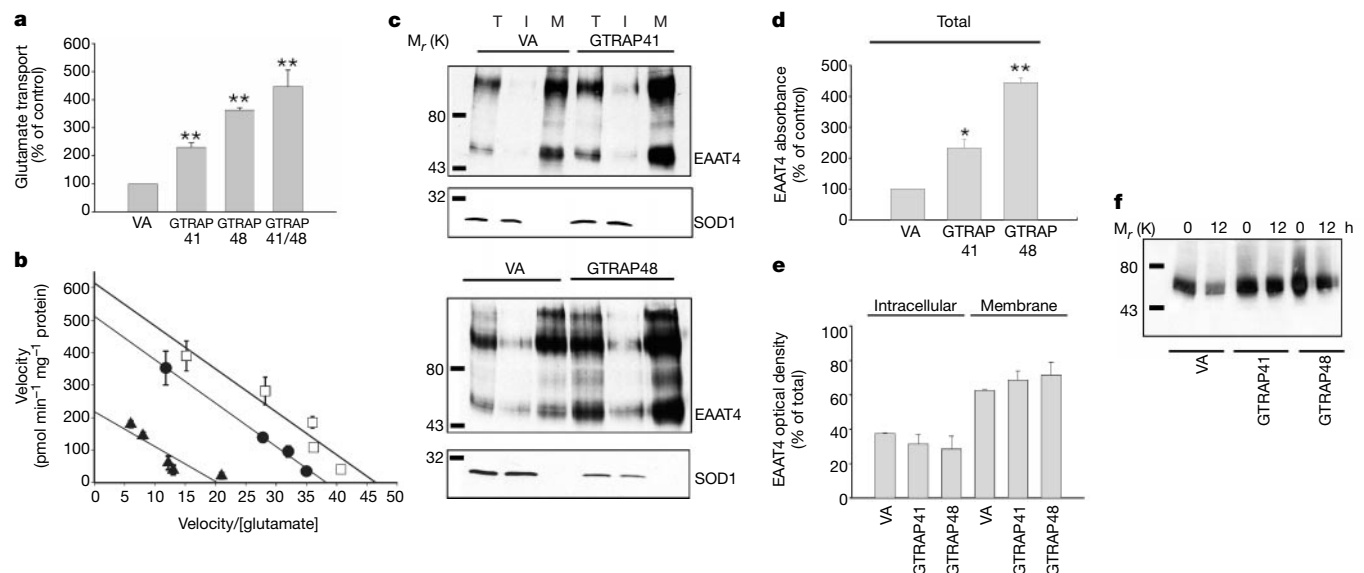


Figure 4 Effect of GTRAP41 and GTRAP48 on Na^+ -dependent [^3H]-glutamate uptake. **a**, GTRAP41 and GTRAP48 expression increased glutamate uptake (5 μM) significantly over cells transfected with vector alone (VA; $n = 4$; $**P < 0.005$). **b**, Kinetic data showed that GTRAP41 (open squares) increased the V_{max} from 222 to 605 $\text{pmol mg}^{-1} \text{min}^{-1}$ and increased the K_m slightly from 7 to 11 μM , as compared with EAAT4 alone (filled triangles). GTRAP48 increased the V_{max} from 208 to 512 $\text{pmol mg}^{-1} \text{min}^{-1}$ (filled circles) and

increased the K_m from 10 to 13 μM . **c**, Immunoblots of total (T), intracellular (I), and biotinylated fractions (M) of HEK-rEAAT4 cells transfected with GTRAP41 and GTRAP48. **d**, Quantitation of immunoblots for total EAAT4 protein ($n = 3$; $**P < 0.005$, $*P < 0.05$). **e**, Ratio of membrane-bound to intracellular EAAT4. **f**, Cells transfected with GTRAP41, GTRAP48 or vector alone incubated with cycloheximide (10 $\mu\text{g ml}^{-1}$) for 12 h.

We determined whether there was a functional association between GTRAP41, GTRAP48 and EAAT4. We measured the sodium-dependent glutamate transport activity of HEK-rEAAT4 cells that had been transfected with one or both of the interacting proteins. GTRAP41 and GTRAP48 produced respective twofold and fourfold increases in glutamate transport (Fig. 4a), and their co-expression resulted in a further increase in glutamate uptake. Kinetic analysis indicated that GTRAP41 and GTRAP48 produced an increase in the $V_{\max}$ of glutamate transport activity (Fig. 4b). GTRAP41 and GTRAP48 may therefore enhance glutamate transport either through an increase in the catalytic rate of the transporter or through an increase in cell-surface availability.

To examine changes in the cell-surface levels of EAAT4, we used a cell-membrane-impermeant biotinylation reagent to label cell-surface proteins selectively. Figure 4c shows western analysis of a representative biotinylation experiment for GTRAP41 and GTRAP48. The total amount of EAAT4 increased when GTRAP41 and GTRAP48 were co-expressed (Fig. 4d); in contrast, the total amount of SOD1, a control for the total amount of protein loaded, was unaltered or decreased in the GTRAP41 or GTRAP48 samples, respectively. The percentage of total EAAT4 that was biotinylated remained the same when GTRAP41 and GTRAP48 were co-expressed, showing that most of the increase in total EAAT4 protein was located at the cell surface and not in an intracellular

pool (Fig. 4e).

These results indicated that GTRAP41 and GTRAP48 stabilize and/or anchor EAAT4 at the cell membrane, making it less likely to be internalized and subsequently degraded; however, we could not rule out the possibility of increased expression of the cell's native gene. To address this issue, we treated cells with cycloheximide, an inhibitor of protein synthesis, 48 h after transfection (Fig. 4f). Densitometry showed that 12 h after treatment the EAAT4 protein in HEK-rEAAT4 cells was reduced to $54 \pm 0.6\%$ of its level before cycloheximide treatment. In contrast, $81 \pm 2\%$ and $74 \pm 1.7\%$ of the EAAT4 protein remained after 12 h when GTRAP41 and GTRAP48 were co-expressed, respectively. These results provide compelling evidence that GTRAP41 and GTRAP48 do stabilize EAAT4 at the membrane.

To determine whether the EAAT4/GTRAP48 interaction is required to mediate the increase in EAAT4 activity, we transfected HEK-rEAAT4 cells with GTRAP48 constructs lacking the last 155 amino acids, which were pulled out by EAAT4 in the yeast two-hybrid screen. The C-terminally truncated GTRAP48 had only a modest effect on stimulating EAAT4 activity, indicating that the protein-protein interaction is responsible for most of the increase in uptake activity (Fig. 5a). We also co-transfected HEK-rEAAT4 cells with GTRAP48 and a Myc-tagged cDNA construct encoding the last 77 amino acids of EAAT4, to disrupt the interaction of GTRAP48 with full-length EAAT4. We found that the co-expression of this construct inhibited the GTRAP48-mediated effect by about 25%, but co-expression of a smaller construct (residues 1,452–1,578), lacking the GTRAP48-binding domain, had no effect (Fig. 5b). Together, these results indicate that the EAAT4/GTRAP48 interaction is important in modulating EAAT4 uptake activity.

The physiological relevance of GTRAP41 and GTRAP48 on EAAT4 uptake activity *in vivo* was subsequently examined by the intra-cisternal injection of HSV amplicon vectors expressing GTRAP41 and GTRAP48. Cerebellar glutamate uptake was measured 48 h after injection, and was elevated when GTRAP41 and GTRAP48 were expressed but not when the control HSVlac amplicon vector was injected (Fig. 5c). Dihydrokainic acid (DHK), an inhibitor of GLT-1-mediated glutamate transport, had no effect on cerebellar glutamate uptake, ruling out any involvement of GLT-1. There is no method to distinguish functionally between GLAST, EAAC1 and EAAT4; but, as GTRAP41 and GTRAP48 do not interact directly with any other transporter, it is likely that the observed increase in uptake is due to an increase in EAAT4 activity. Western blot analysis confirmed an increased expression of GTRAP41 and GTRAP48 in the cerebellum after the injections. Preliminary studies of primary cultures of rat Purkinje cells indicate that EAAT4 and GTRAP41 may colocalize perisynaptically at 70% of synapses.

Glutamate transporters, through their rapid and efficient removal of glutamate from the synaptic cleft, are critical in glutamatergic plasticity and the prevention of glutamate-mediated excitotoxicity. Identifying and characterizing the glutamate transporter associated proteins GTRAP41 and GTRAP48 has begun to unravel the complex mechanism of glutamate uptake. In the modulation of EAAT4 glutamate transport, our findings implicate a role for G-protein signalling, a pathway that may involve Rho activation, and anchoring to the actin cytoskeleton (Fig. 5d). These proteins may also modulate the perisynaptic distribution of EAAT4 at glutamatergic synapses. Future studies are required to delineate the signalling pathway of GTRAP41 and GTRAP48, and how their interaction may be relevant to normal and abnormal glutamatergic neurotransmission, as altered EAAT4 function may be involved in the neurodegenerative disease spinocerebellar ataxia (SCA1)²⁰. By understanding the physiological regulation of the EAAT4 transporter, we may be able to identify possible therapies for SCA1 or other toxic insults that lead to the degeneration of Purkinje cell neurons.

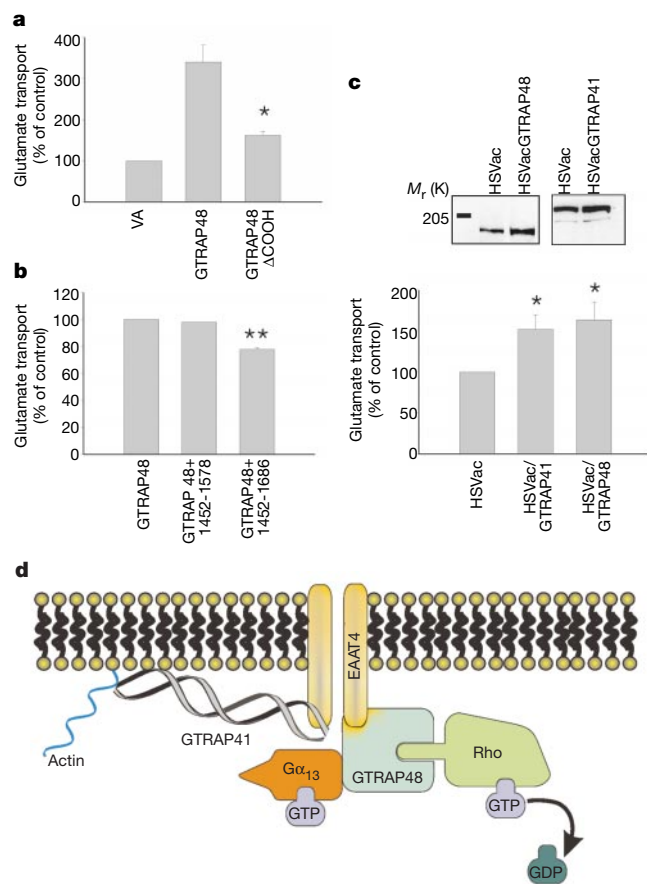


Figure 5 Modulation of EAAT4 activity. **a**, GTRAP48 Δ COOH results in a very small increase in EAAT4 uptake activity compared with full-length GTRAP48 ($*P < 0.05$). **b**, Overexpression of the EAAT4 C-terminus reduced the GTRAP48-mediated effect on EAAT4 activity ($**P < 0.005$). **c**, Intra-cisternal injection of HSVlac amplicons expressing GTRAP41 and GTRAP48 increased glutamate uptake *in vivo* ($n = 6$; $*P < 0.05$). Western blot analysis demonstrates increased expression of GTRAP41 and GTRAP48. **d**, A possible model of the coupling of EAAT4 to a Rho GTPase signal transduction cascade and to the actin cytoskeleton via GTRAP48 and GTRAP41. GDP, guanosine diphosphate; GTP, guanosine triphosphate.

Methods

Yeast two-hybrid screen

We screened a rat brain cDNA library (Clontech) using the final 77 amino acids of EAAT4 as bait (pGBT9). For the EAAT4 C-terminal domain analysis, different regions of the final 77 amino acids of EAAT4 were subcloned in-frame into the pGBT9 vector.

Cloning of full-length GTRAP41 and GTRAP48 cDNAs

Marathon cDNA amplification (CLONTECH) was used to perform both 5' and 3'-RACE on cDNA synthesized from rat brain poly(A)⁺ RNA.

Antibodies

Affinity-purified polyclonal antisera to EAAT4, GTRAP41 and GTRAP48 were generated as described⁶. We synthesized peptides corresponding to epitopes of EAAT4 (C-terminal, EKGASRGGRGNESA; N-terminal, KNSLFLRESGAGGGCL), rat GTRAP41 (KRGPAPSPMPQSRSSSE) and rat GTRAP48 (KTPERTSPSHHRQPSD). The anti-c-Myc monoclonal antibody was from Boehringer Mannheim. For visualization of intracellular F-actin organization, the cells were probed with Texas-red-conjugated or fluorescein isothiocyanate (FITC)-conjugated phalloidin (Sigma).

Transfection and maintenance of HEK-rEAAT4 cells

The EAAT4 cDNA was subcloned into pcDNA3.1/Hygro(+) (Invitrogen) using the *EcoRI* restriction site. The plasmid was linearized with *SspI* and transfected into HEK 293T cells. Forty-eight hours after transfection, the cells were split to 50% confluency, and hygromycin (Invitrogen) was added at a concentration of 50 µg ml⁻¹. After about 2–3 weeks of selection, a serial dilution was carried out and cells were plated out, without selection, in a 96-well plate to obtain one cell per well. Several colonies were picked, expanded in selective medium and checked for expression by western blotting.

Fusion proteins and *in vitro* binding

Full-length EAAT4 was subcloned into the *EcoRI* site of the GST–fusion vector pGEX-6P-1 (Pharmacia). Synthesis of recombinant proteins in BL21 cells (Novagen) was induced by 0.1 mM isopropyl β-D-thiogalactoside for 2 h at 30 °C and purified according to the protocol provided by Pharmacia. HEK 293T cells were transfected with Myc-tagged GTRAP41 or GTRAP48 and gathered in ice-cold immunoprecipitation (IP) buffer. The cellular lysate was incubated with GST or GST–EAAT4 immobilized on glutathione–Sephadex 4B, and washed to remove nonspecifically bound proteins. Specifically bound proteins were eluted with 2× SDS loading buffer and analysed by immunoblotting using an anti-c-Myc antibody.

Co-immunoprecipitation

Full-length GTRAP41 and GTRAP48 cDNAs were subcloned into the *NotI* site of a Myc-tagged pRK5 vector and used to transfect HEK-rEAAT4 cells. After transfection (48–72 h), cells were solubilized with 1 ml of ice-cold IP buffer for 2 h at 4 °C with rotation and centrifuged to remove cellular debris. Antibody was added to 0.5 ml of supernatant and incubated overnight at 4 °C. We dissected and prepared the cerebellum from a Sprague-Dawley rat as described²¹. For each immunoprecipitation, 500 µg of the Triton-lysate was incubated overnight at 4 °C with 5 µg of antibody. Protein-A–Sephadex (Pharmacia) was then added for 2 h at 4 °C, washed once with IP buffer and three times with IP buffer minus Triton X-100. Bound protein was eluted by boiling in 2× SDS loading buffer, and analysed by immunoblotting.

Immunohistochemistry

Rat brain sections were stained as described⁹, using the following antibodies: C-terminal anti-EAAT4 (1.5 µg ml⁻¹), anti-GTRAP41 (127 ng ml⁻¹) or anti-GTRAP48 (132 ng ml⁻¹) antibodies. Texas-red and FITC-conjugated secondary antibodies were used at dilutions of 1:200.

Guanine nucleotide exchange assay

Small G proteins GST–RhoA, GST–CDC42 and GST–Rac were expressed in bacterial cells and affinity purified to ~80% purity using a glutathione column. We incubated 20 pmoles of each protein with 100 pmoles GTP-γS for 10 min at 30 °C with varying concentrations of full-length GTRAP48 or p115. The binding reactions were filtered through BA-85 nitrocellulose and the amount of GTP-γS bound to small G protein was quantified by scintillation counting of the dried filters. Binding of Gα₁₃ to GTRAP48 was assayed as described¹⁸.

Na⁺-dependent glutamate transport activity

HEK-rEAAT4 cells transfected with GTRAP41 and GTRAP48 were grown in a monolayer on six-well plates, and assays were conducted 72 h after transfection as described²². We subcloned GTRAP41 and GTRAP48 into the *EcoRI* site of HSVPrPUC amplicon parent vector²³. Amplicon vector DNA (3.6 µg) and pBAC-V2 DNA (25 µg) were used to transfect 2 × 10⁷ baby hamster kidney cells as described²⁴. Virus was collected 72 h after transfection, and titred as described²⁵. Expression particles (2 × 10⁷) were injected intracisternally into male Sprague-Dawley rats (250 g) obtained from Zivic Miller. Forty-eight hours after injection the rats were killed, synaptosomal preparations of the cerebelli were prepared using a polytron, and glutamate uptake was measured.

Biotinylation

Biotinylation of cell-surface proteins was done as described²⁶. We used SOD1 to control for total protein and to determine whether the biotinylation reagent labels proteins in the intracellular compartment.

Statistics

Statistical differences were determined by Students's *t*-test for two-group comparisons.

Received 28 April; accepted 18 December 2000.

1. Storck, T., Schulte, S., Hofmann, K. & Stoffel, W. Structure, expression, and functional analysis of a Na⁺-dependent glutamate/aspartate transporter from rat brain. *Proc. Natl Acad. Sci. USA* **89**, 10955–10959 (1992).
2. Kanai, Y. & Hediger, M. A. Primary structure and functional characterization of a high-affinity glutamate transporter. *Nature* **360**, 467–471 (1992).
3. Pines, G. *et al.* Cloning and expression of a rat brain L-glutamate transporter. *Nature* **360**, 464–467 (1992); erratum *ibid.* **360**, 768 (1992).
4. Fairman, W. A., Vandenberg, R. J., Arriza, J. L., Kavanaugh, M. P. & Amara, S. G. An excitatory amino acid transporter with properties of a ligand-gated chloride channel. *Nature* **375**, 599–603 (1995).
5. Arriza, J. L., Eliasof, S., Kavanaugh, M. P. & Amara, S. G. Excitatory amino acid transporter 5, a retinal glutamate transporter coupled to a chloride conductance. *Proc. Natl Acad. Sci. USA* **94**, 4155–4160 (1997).
6. Rothstein, J. D. *et al.* Localization of neuronal and glial glutamate transporters. *Neuron* **13**, 713–725 (1994).
7. Danbolt, N. C., Storm-Mathisen, J. & Kanner, B. I. An [Na⁺ + K⁺] coupled L-glutamate transporter purified from rat brain is located in glial cell processes. *Neuroscience* **51**, 295–310 (1992).
8. He, Y., Janssen, W. G., Rothstein, J. D. & Morrison, J. H. Differential synaptic localization of the glutamate transporter EAAC1 and glutamate receptor subunit GluR2 in the rat hippocampus. *J. Comp. Neurol.* **418**, 255–269 (2000).
9. Furuta, A., Martin, L. J., Lin, C. L., Dykes-Hoberg, M. & Rothstein, J. D. Cellular and synaptic localization of the neuronal glutamate transporters excitatory amino acid transporter 3 and 4. *Neuroscience* **81**, 1031–1042 (1997).
10. Dehnes, Y. *et al.* The glutamate transporter EAAT4 in rat cerebellar Purkinje cells: a glutamate-gated chloride channel concentrated near the synapse in parts of the dendritic membrane facing astroglia. *J. Neurosci.* **18**, 3606–3619 (1998).
11. Nagao, S., Kwak, S. & Kanazawa, I. EAAT4, a glutamate transporter with properties of a chloride channel, is predominantly localized in Purkinje cell dendrites, and forms parasagittal compartments in rat cerebellum. *Neuroscience* **78**, 929–933 (1997).
12. Tong, G. & Jahr, C. E. Block of glutamate transporters potentiates postsynaptic excitation. *Neuron* **13**, 1195–1203 (1994).
13. Diamond, J. S. & Jahr, C. E. Transporters buffer synaptically released glutamate on a submillisecond time scale. *J. Neurosci.* **17**, 4672–4687 (1997).
14. Overstreet, L. S., Kinney, G. A., Liu, Y. B., Billups, D. & Slater, N. T. Glutamate transporters contribute to the time course of synaptic transmission in cerebellar granule cells. *J. Neurosci.* **19**, 9663–9673 (1999).
15. Barbour, B., Keller, B. U., Llano, I. & Marty, A. Prolonged presence of glutamate during excitatory synaptic transmission to cerebellar Purkinje cells. *Neuron* **12**, 1331–1343 (1994).
16. Hart, M. J. *et al.* Identification of a novel guanine nucleotide exchange factor for the Rho GTPase. *J. Biol. Chem.* **271**, 25452–25458 (1996).
17. Fukuhara, S., Murga, C., Zohar, M., Igishi, T. & Gutkind, J. S. A novel PDZ domain containing guanine nucleotide exchange factor links heterotrimeric G proteins to Rho. *J. Biol. Chem.* **274**, 5868–5879 (1999).
18. Hart, M. J. *et al.* Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by Gα₁₃. *Science* **280**, 2112–2114 (1998).
19. Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **279**, 509–514 (1998).
20. Lin, X., Antalffy, B., Kang, D., Orr, H. T. & Zoghbi, H. Y. Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1. *Nature Neurosci.* **3**, 157–163 (2000).
21. Blackstone, C. D. *et al.* Biochemical characterization and localization of a non-N-methyl-D-aspartate glutamate receptor in rat brain. *J. Neurochem.* **58**, 1118–1126 (1992).
22. Davis, K. E. *et al.* Multiple signaling pathways regulate cell surface expression and activity of the excitatory amino acid carrier 1 subtype of Glu transporter in C6 glioma. *J. Neurosci.* **18**, 2475–2485 (1998).
23. Geller, A. I., Keyomarsi, K., Bryan, J. & Pardee, A. B. An efficient deletion mutant packaging system for defective herpes simplex virus vectors: potential applications to human gene therapy and neuronal physiology. *Proc. Natl Acad. Sci. USA* **87**, 8950–8954 (1990).
24. Stavropoulos, T. A. & Strathdee, C. A. An enhanced packaging system for helper-dependent herpes simplex virus vectors. *J. Virol.* **72**, 7137–7143 (1998).
25. Bowers, W. J., Howard, D. F. & Federoff, H. J. Discordance between expression and genome transfer titrating of HSV amplicon vectors: recommendation for standardized enumeration. *Mol. Ther.* **1**, 294–299 (2000).
26. Duan, S., Anderson, C. M., Stein, B. A. & Swanson, R. A. Glutamate induces rapid upregulation of astrocyte glutamate transport and cell-surface expression of GLAST. *J. Neurosci.* **19**, 10193–10200 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. J. Maragakis, M. Watanabe, A. Sawa, R. Ganel, J. Llado and R. Law for discussions, advice and help. We thank R. Haganir for the pRK5 vector and D. Howard for technical assistance. This work was supported by the NIH.

Correspondence and requests for materials should be addressed to J.D.R. (e-mail: jrothste@welchlink.welch.jhu.edu). The GenBank accession numbers are AF225960 (GTRAP41) and AF225961 (GTRAP48).

Aberrant CFTR-dependent HCO_3^- transport in mutations associated with cystic fibrosis

Joo Young Choi*, Daniella Muallem*, Kirill Kiselyov*, Min Goo Lee †, Philip J. Thomas* & Shmuel Muallem*

* Department of Physiology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

† Department of Pharmacology and Brain Korea 21 Project for Medical Sciences, Yonsei University College of Medicine, Seoul 120-752, Korea

Cystic fibrosis (CF) is a disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). Initially, Cl^- conductance in the sweat duct was discovered to be impaired in CF¹, a finding that has been extended to all CFTR-expressing cells^{2–4}. Subsequent cloning of the gene^{5,6} showed that CFTR functions as a cyclic-AMP-regulated Cl^- channel⁷; and some CF-causing mutations inhibit CFTR Cl^- channel activity^{2–4,8}. The identification of additional CF-causing mutants with normal Cl^- channel activity indicates, however, that other CFTR-dependent processes contribute to the disease. Indeed, CFTR regulates other transporters^{3,4}, including Cl^- -coupled HCO_3^- transport^{9,10}. Alkaline fluids are secreted by normal tissues, whereas acidic fluids are secreted by mutant CFTR-expressing tissues¹¹, indicating the importance of this activity. HCO_3^- and pH affect mucin viscosity^{12,13} and bacterial binding^{14,15}. We have examined Cl^- -coupled HCO_3^- transport by CFTR mutants that retain substantial or normal Cl^- channel activity. Here we show that mutants reported to be associated with CF with pancreatic insufficiency

do not support HCO_3^- transport, and those associated with pancreatic sufficiency show reduced HCO_3^- transport. Our findings demonstrate the importance of HCO_3^- transport in the function of secretory epithelia and in CF.

We compared the activity of wild-type CFTR to that of 17 disease-causing mutants. Several criteria were used for selection of mutants and the methods to evaluate Cl^- and HCO_3^- transport. All mutants selected code for properly processed proteins^{16–21}. The macroscopic Cl^- current and single-channel activity of the mutants have been shown to be significant, normal or even elevated^{16–21}. The mutants are from all cytoplasmic domains of CFTR and have been reported to be associated with CF with pancreatic sufficiency or pancreatic insufficiency. Table 1 in the Supplementary Information lists the mutants and the reported clinical status of the patients. Two of the mutations, I148T and R117H, are relatively common, and thus the clinical data are solid. Initially we analysed these mutations to establish the relationship between the CF phenotype and HCO_3^- transport.

CFTR-dependent HCO_3^- transport is dependent on Cl^- transport and is not affected by the membrane potential^{9,10}, suggesting that it is an electroneutral process that can only be followed by measuring intracellular HCO_3^- concentration. Hence, changes in HCO_3^- concentration were evaluated from changes in intracellular pH (pH_i) (see refs 9, 10). To evaluate Cl^- transport under the same conditions, we monitored intracellular chloride concentrations ($[\text{Cl}^-]_i$) with *N*-(6-methoxyquinolyl)acetoethyl ester (MQAE). To validate the ability of this procedure to accurately report Cl^- channel activity, we determined the correlation between expression efficiency (green fluorescent protein (GFP) fluorescence), Cl^- current and changes in $[\text{Cl}^-]_i$ in cells transfected with CFTR, CFTR(I148T) and CFTR(R117H). Figure 1a–h shows that similar results are reported by measurement of Cl^- current and $[\text{Cl}^-]_i$ for CFTR and the mutants. Therefore, the Cl^- transport capacity of all other CFTR

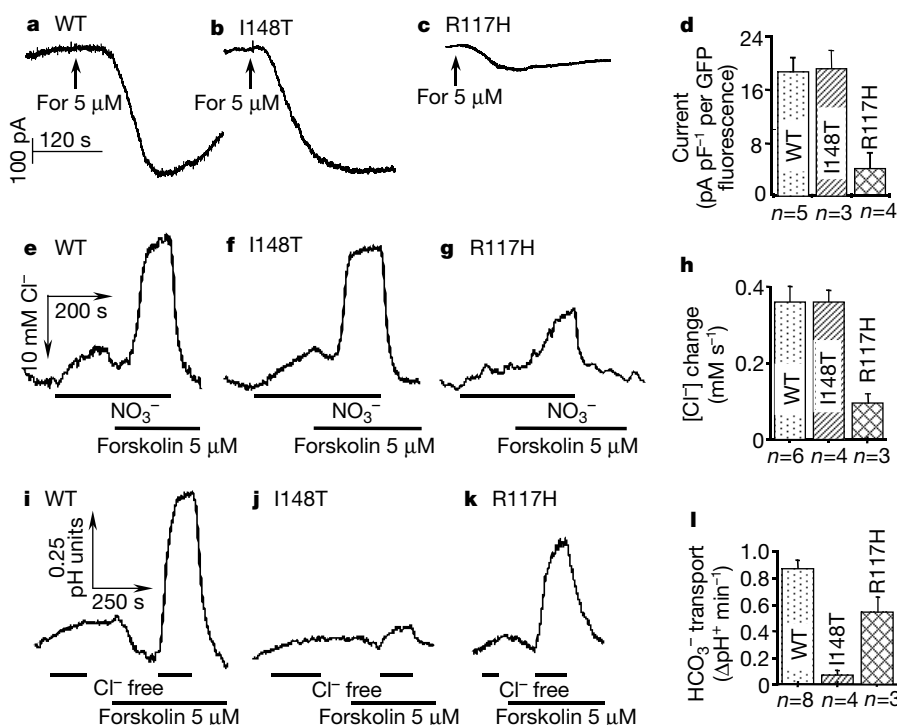


Figure 1 cAMP-stimulated Cl^- and HCO_3^- transport by wild-type (WT) CFTR and the CFTR mutants I148T and R117H. The whole-cell Cl^- current of HEK293 cells transfected with the indicated constructs was determined (a–c). For, forskolin. The current was normalized with respect to membrane capacitance and GFP fluorescence before averaging (d). Transfected cells were also loaded with MQAE (e–h) or BCECF (i–l) for

measurements of $[\text{Cl}^-]_i$ and pH_i , respectively. Cells loaded with MQAE were exposed to a solution in which Cl^- was replaced with NO_3^- and then stimulated with 5 μM forskolin. For pH_i measurements, Cl^- was replaced with gluconate. After calibration, initial rates of changes in $[\text{Cl}^-]_i$ (h) and pH_i (i) were averaged.

mutants was evaluated from changes in $[Cl^-]_i$.

Figure 1e–l shows the experimental protocols used to measure the effect of the I148T and R117H mutations on Cl^- and HCO_3^- transport. The I148T mutant is normally processed and mediates macroscopic Cl^- current (Fig. 1d; ref. 16), single-channel properties¹⁶ and Cl^- fluxes (Fig. 1h) indistinguishable from those of wild-type CFTR, but is associated with CF with pancreatic insufficiency. Replacing external Cl^- with NO_3^- in non-stimulated cells caused a slow Cl^- efflux in cells expressing CFTR or the I148T mutant. Stimulation of CFTR-expressing cells with forskolin resulted in a change in $[Cl^-]_i$ at a rate of 0.36 ± 0.04 ($n = 6$) $mM s^{-1}$. Similar rates were measured for the I148T, G178R, A1067T, G1244E, S1255P and G1349D mutants (see Fig. 3 for location of these mutations in CFTR), all of which are associated with CF with pancreatic insufficiency. The rates of changes in $[Cl^-]_i$ measured here, and the macroscopic Cl^- currents from previous studies are listed in Table 1 in the Supplementary Information. In all cases, the presence of forskolin-stimulated Cl^- fluxes confirmed that the CFTR mutants matured and travelled properly to the plasma membrane.

The similarity of Cl^- transport by CFTR and the mutants implies that aberrant Cl^- transport cannot alone account for these severe forms of CF. Stimulation of CFTR markedly increases Cl^- -coupled HCO_3^- transport (Fig. 1i) by an unknown mechanism. Several studies have reported that CFTR functions as both a Cl^- and a HCO_3^- channel^{22–24}, which might account for the fluxes reported here. In a continuing work, however, we confirmed that Cl^- efflux is obligatory for CFTR-dependent HCO_3^- influx, and that depolarizing the cells does not inhibit the transport by CFTR and the mutants. Therefore, we consider the CFTR-dependent transport as Cl^- -coupled HCO_3^- transport. Remarkably, all mutations associated with CF with pancreatic insufficiency tested here, without exception, eliminate the ability of CFTR to support HCO_3^- transport. Results for the I148T mutant are shown in Fig. 1, and the averaged results for all mutants are listed in Table 1 in the Supplementary Information.

We next analysed Cl^- and HCO_3^- transport in CFTR mutants associated with pancreatic sufficiency. The results obtained with the R117H mutation are illustrated in Fig. 1. In sharp contrast with the results obtained with the mutants associated with CF with pancreatic insufficiency, and in agreement with previous reports²⁵, the R117H mutation reduced Cl^- current and the MQAE response by about 70%. By contrast, the R117H mutation reduced the ability of

CFTR to support HCO_3^- transport by only 37%. The E193K, D648V, H949Y and R1070Q mutants, all associated with CF with pancreatic sufficiency, had no effect on Cl^- transport but reduced HCO_3^- transport by 50–65%. Two sets of particularly interesting mutants are G551D and G551S and H620Q and A800G. G551S, which is associated with CF with pancreatic sufficiency, had no effect on Cl^- transport but reduced HCO_3^- transport by 59% (Fig. 2). As reported previously^{17,18,26}, G551D reduced Cl^- transport to 53% of the control value. By contrast, this mutation, which is associated with CF with pancreatic insufficiency, completely inhibited CFTR-dependent HCO_3^- transport (Fig. 2b, d).

When expressed in oocytes, the H620Q and A800G mutants increased the macroscopic Cl^- current about threefold and the channel open probability by 150–180% (ref. 19). We confirmed these findings, as Cl^- transport by these mutants is about 2.4-fold higher than that by CFTR (Fig. 2e, f). Despite these markedly enhanced Cl^- fluxes, the A800G mutation, which causes CF with pancreatic sufficiency, only stimulated HCO_3^- transport to 75% of the wild-type CFTR value. The H620Q mutant, found in one CF patient with pancreatic insufficiency, only stimulated HCO_3^- transport about 13% as effectively as did CFTR. The enhanced Cl^- transport by the two mutants in heterologous systems seems to be caused by increased expression of the protein in the plasma membrane¹⁹. If this enhanced expression is not maintained *in vivo*, the mutations will further compromise HCO_3^- transport to exacerbate the pancreatic phenotypes. In this respect, the R1070Q mutation, in the few documented cases, has been found to be associated with CF with either pancreatic sufficiency or pancreatic insufficiency. Our results of substantial HCO_3^- transport by this mutant would predict a phenotype of CF with pancreatic sufficiency.

To account for variable Cl^- transport and possible variable expression of the proteins in the plasma membrane, we normalized the capacity of the CFTR mutants to transport HCO_3^- with respect to their capacity to transport Cl^- . Figure 3 shows the correlation between the reported pancreatic status of the patients and the $HCO_3^-:Cl^-$ transport ratio. When the ratio measured with CFTR is taken as 1, all mutants associated with CF with pancreatic insufficiency show an $HCO_3^-:Cl^-$ transport ratio of less than 0.1. By comparison, all mutations associated with CF with pancreatic sufficiency show an $HCO_3^-:Cl^-$ transport ratio of between 0.31 and 0.46. Notably, although a few of these mutants exhibit altered

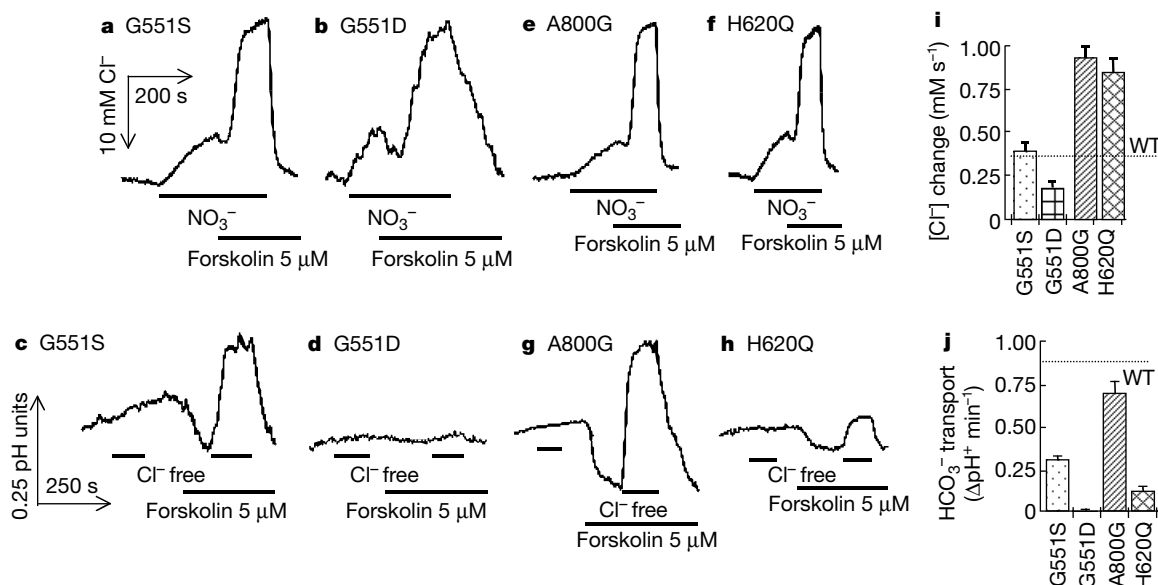


Figure 2 cAMP-stimulated Cl^- and HCO_3^- transport by CFTR mutants associated with a severe or a mild form of CF. $[Cl^-]_i$ (a, b, e, f) and HCO_3^- (c, d, g, h) transport were measured as in Fig. 1. Summaries of averaged rates are given for Cl^- (i) and for HCO_3^- (j).

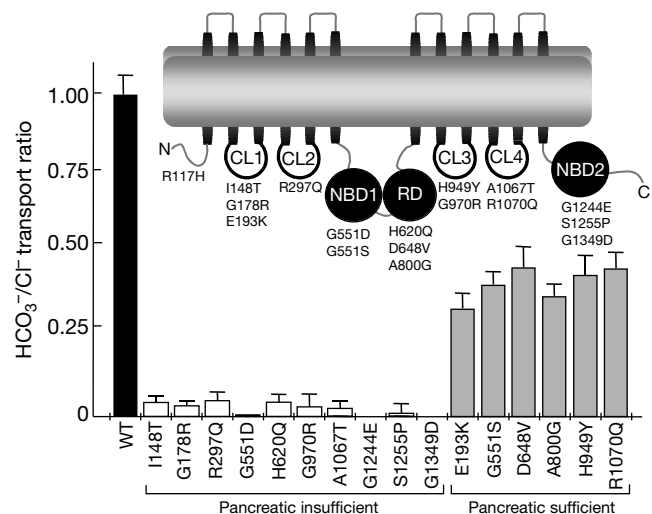


Figure 3 The HCO₃⁻:Cl⁻ transport ratio of CFTR mutants associated with CF. The HCO₃⁻:Cl⁻ transport ratios were calculated from the averaged rates summarized in Table 1 of the Supplementary Information. The ratio measured in wild-type CFTR was set to 1. Inset illustrates the different cytoplasmic domains of CFTR. CL1, cytoplasmic loop 1; CL2, cytoplasmic loop 2; NBD1, nucleotide-binding domain 1; RD, regulatory domain; CL3, cytoplasmic loop 3; CL4, cytoplasmic loop 4; NBD2, nucleotide-binding domain 2.

channel gating or reduced processing, in all of the mutants CFTR-dependent HCO₃⁻ transport is affected more than CFTR-mediated Cl⁻ transport. In this regard, although the R117H mutation markedly reduces Cl⁻ channel activity and Cl⁻ transport (Fig. 1), it is associated with a mild form of CF, probably because it supports substantial HCO₃⁻ transport (Fig. 1). Consequently, the CFTR-dependent HCO₃⁻:Cl⁻ transport ratio seems to correlate well with the reported pancreatic function of CF patients.

Cl⁻ transport across the apical membrane of secretory epithelia is the rate-limiting step in fluid and electrolyte transport^{1-4,8}, highlighting the importance of the Cl⁻ channel function of CFTR. The existence of CF-causing CFTR mutations that support normal or even elevated Cl⁻ transport indicates, however, that factors other than abnormal Cl⁻ transport can also lead to CF. An alternative view of the role of CFTR in epithelial transport was introduced with the finding that CFTR can affect the transport of other ions²⁷⁻³⁰, including HCO₃⁻ (refs 9, 10, 22-24). The aberrant HCO₃⁻ transport by the CF-causing mutations examined here indicates that HCO₃⁻ transport by CFTR-expressing epithelia is critical for normal tissue physiology, and that impaired HCO₃⁻ transport is sufficient to derange pancreatic function even in the presence of Cl⁻ channel activity. Acidic fluid secretion by CFTR-expressing tissues in the disease state may lead to precipitation of mucins and plugging of ductal systems^{12,13}, and facilitate bacterial infection through binding to the precipitated mucins^{14,15}. In the special case of the pancreas, acidic pH would lead to premature activation of digestive enzymes, destruction of the pancreas and pancreatic insufficiency⁸. Thus, the aberrant HCO₃⁻ transport model can account for diverse pathologies observed in the disease. Our findings suggest that simple correction of Cl⁻ transport is not likely to ameliorate the symptoms of CF. Enhancing HCO₃⁻ transport by epithelial cells, even in the absence of CFTR, or increasing the HCO₃⁻ content on the apical surface of affected tissues, should be considered as additional means of reducing the debilitating effects of CF. □

Methods

Site-directed mutagenesis

The pCMVNot 6.2 plasmid carrying the human wild-type CFTR gene was a gift from J. Rommens (Hospital for Sick Children, Toronto, Canada). Site-directed mutagenesis was performed using the QuickChange site-directed mutagenesis kit from Stratagene (La Jolla,

CA). All mutations were verified by sequencing of four separate clones from each mutant. The activity of all four clones was measured to verify that all have similar activity.

Expression of CFTR and cell transfection

All wild-type and CFTR mutants were expressed in HEK293 cells by transient transfection using the lipofectamine reagent. To control for expression of CFTR and identify the CFTR-expressing cells, we co-transfected the cells with a GFP-expressing plasmid (Life Technologies). In previous work, we showed that there is good agreement between expression of GFP and CFTR⁹. GFP fluorescence was measured before current recording or loading the cells with 2',7'-bis-(2-carboxyethyl)-5-(6)-carboxyfluorescein (BCECF) or after loading with MQAE. Hence, all experiments were performed with cells expressing comparable amount of GFP.

Cl⁻ currents, [Cl⁻]_i, and pH_i

HCO₃⁻ and Cl⁻ transport were measured with matched pairs of cells transfected in the same dish. In brief, cells grown on two cover slips in each dish were transfected with the same solution and maintained in culture together until use 48-72 h after transfection. On the day of experiment, cells on one of the cover slips were loaded with BCECF (15 min at room temperature) and used immediately to measure HCO₃⁻ transport. At the same time, cells on the matching cover slip were loaded with MQAE by 1.5-2 h at 37 °C in culture medium containing 5 mM MQAE and then used to measure [Cl⁻]_i.

The procedures and solutions used to measure the CFTR-mediated, whole-cell Cl⁻ current in the HEK 293 cells transfected with wild-type CFTR and the CFTR mutants were identical to those described for measurement of CFTR-mediated Cl⁻ current in NIH 3T3 cells⁹. Measurement of pH_i and [Cl⁻]_i with BCECF and MQAE techniques, respectively, in single HEK293 cells, calibration of fluorescent signals and composition of solutions have been described^{9,10}.

Received 2 November; accepted 11 December 2000.

- Quinton, P. M. Chloride impermeability in cystic fibrosis. *Nature* **301**, 421-422 (1983).
- Quinton, P. M. Physiological basis of cystic fibrosis: a historical perspective. *Physiol. Rev.* **79**, S3-S22 (1999).
- Pilewski, J. M. & Frizzell, R. A. Role of CFTR in airway disease. *Physiol. Rev.* **79**, S215-S255 (1999).
- Grubb, B. R. & Boucher, R. C. Pathophysiology of gene-targeted mouse models for cystic fibrosis. *Physiol. Rev.* **79**, S193-S213 (1999).
- Rommens, J. M. *et al.* Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-1065 (1989).
- Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066-1073 (1989).
- Sheppard, D. N. & Welsh, M. J. Structure and function of the CFTR chloride channel. *Physiol. Rev.* **79**, S23-S45 (1999).
- Argent, B. E. & Case, R. M. in *Physiology of the Gastrointestinal Tract* 3rd edn (ed. Johnson, L. R.) 1478-1498 (Raven, New York, 1994).
- Lee, M. G. *et al.* Regulation of Cl⁻/HCO₃⁻ exchange by cystic fibrosis transmembrane conductance regulator expressed in NIH 3T3 and HEK 293 cells. *J. Biol. Chem.* **274**, 3414-3421 (1999).
- Lee, M. G. *et al.* Cystic fibrosis transmembrane conductance regulator regulates luminal Cl⁻/HCO₃⁻ exchange in mouse submandibular and pancreatic ducts. *J. Biol. Chem.* **274**, 14670-14677 (1999).
- Kopelman, H. *et al.* Impaired chloride secretion, as well as bicarbonate secretion, underlies the fluid secretory defect in the cystic fibrosis pancreas. *Gastroenterology* **95**, 349-355 (1988).
- Veerman, E. C., Valentijn-Benz, M. & Nieuw Amerongen, A. V. Viscosity of human salivary mucins: effect of pH and ionic strength and role of sialic acid. *J. Biol. Buccale* **17**, 297-306 (1989).
- Bhaskar, K. R. *et al.* Profound increase in viscosity and aggregation of pig gastric mucin at low pH. *Am. J. Physiol.* **261**, G827-G832 (1991).
- Wanke, C. A., Cronan, S., Goss, C., Chadee, K. & Guerrant, R. L. Characterization of binding of *Escherichia coli* strains which are enteropathogens to small-bowel mucin. *Infect. Immun.* **58**, 794-800 (1990).
- Veerman, E. C. *et al.* Sulfated glycans on oral mucin as receptors for *Helicobacter pylori*. *Glycobiology* **7**, 737-743 (1997).
- Seibert, F. S. *et al.* Disease-associated mutations in cytoplasmic loops 1 and 2 of cystic fibrosis transmembrane conductance regulator impede processing or opening of the channel. *Biochemistry* **36**, 11966-11974 (1997).
- Illek, B. *et al.* Defective function of the cystic fibrosis-causing missense mutation G551D is recovered by genistein. *Am. J. Physiol.* **277**, C833-C839 (1999).
- Anderson, M. & Welsh, M. J. Regulation by ATP and ADP of CFTR chloride channels that contain mutant nucleotide-binding domains. *Science* **257**, 1701-1704 (1992).
- Vankeerberghen, A. *et al.* Characterization of 19 disease-associated missense mutations in the regulatory domain of the cystic fibrosis transmembrane conductance regulator. *Hum. Mol. Genet.* **7**, 1761-1769 (1998).
- Seibert, F. S. *et al.* Cytoplasmic loop three of cystic fibrosis transmembrane conductance regulator contributes to regulation of chloride channel activity. *J. Biol. Chem.* **271**, 27493-27499 (1996).
- Seibert, F. S. *et al.* Disease-associated mutations in the fourth cytoplasmic loop of cystic fibrosis transmembrane conductance regulator compromise biosynthetic processing and chloride channel activity. *J. Biol. Chem.* **271**, 15139-15145 (1996).
- Gray, M. A., Plant, S. & Argent, B. E. cAMP-regulated whole cell chloride currents in pancreatic duct cells. *Am. J. Physiol.* **264**, C591-C602 (1993).
- Poulsen, J. H., Fischer, H., Illek, B. & Machen, T. E. Bicarbonate conductance and pH regulatory capability of cystic fibrosis transmembrane conductance regulator. *Proc. Natl. Acad. Sci. USA* **91**, 5340-5344 (1994).
- Linsdell, P. *et al.* Permeability of wild-type and mutant cystic fibrosis transmembrane conductance regulator chloride channels to polyatomic anions. *J. Gen. Physiol.* **110**, 355-364 (1997).
- Sheppard, D. N. *et al.* Mutations in CFTR associated with mild-disease-form Cl⁻ channels with altered pore properties. *Nature* **362**, 160-164 (1993).

26. Fulmer, S. B., Schwiebert, E. M., Morales, M. M., Guggino, W. B. & Cutting, G. R. Two cystic fibrosis transmembrane conductance regulator mutations have different effects on both pulmonary phenotype and regulation of outwardly rectified chloride currents. *Proc. Natl Acad. Sci. USA* **92**, 6832–6836 (1995).
27. Stutts, M. J. *et al.* CFTR as a cAMP-dependent regulator of sodium channels. *Science* **269**, 847–850 (1995).
28. Mall, M., Bleich, M., Greger, R., Schreiber, R. & Kunzelmann, K. The amiloride-inhibitable Na^+ conductance is reduced by the cystic fibrosis transmembrane conductance regulator in normal but not in cystic fibrosis airways. *J. Clin. Invest.* **102**, 15–21 (1998).
29. Reddy, M. M., Light, M. J. & Quinton, P. M. Activation of the epithelial Na^+ channel (ENaC) requires CFTR Cl^- channel function. *Nature* **402**, 301–304 (1999).
30. McNicholas, C. M. *et al.* Sensitivity of a renal K^+ channel (ROMK2) to the inhibitory sulfonylurea compound glibenclamide is enhanced by coexpression with the ATP-binding cassette transporter cystic fibrosis transmembrane regulator. *Proc. Natl Acad. Sci. USA* **93**, 8083–8088 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank H. Cuppens from the European CF Consortium for informing us that the G970R mutation results in pancreatic insufficiency. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to S.M. (e-mail: shmuel.muallem@email.swmed.edu).

Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice

Elizabeth A. Lindsay*, Francesca Vitelli*, Hong Su†, Masae Morishima*, Tuong Huynh*, Tiziano Pramparo*, Vesna Jurecic‡, George Ogunrinu§, Helen F. Sutherland||, Peter J. Scambler||, Allan Bradley†‡§# & Antonio Baldini*†

* Departments of Pediatrics (Cardiology) and † Molecular & Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

‡ University of Miami School of Medicine, 1600NW 10th Avenue, Miami, Florida 33136, USA

§ Department of Molecular Genetics, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA

|| Institute of Child Health, 30 Guilford Street, London WC1N 1UK

¶ Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1S, UK

Howard Hughes Medical Institute, One Baylor Plaza, Houston, Texas 77030, USA

DiGeorge syndrome is characterized by cardiovascular, thymus and parathyroid defects and craniofacial anomalies, and is usually caused by a heterozygous deletion of chromosomal region 22q11.2 (*del22q11*) (ref. 1). A targeted, heterozygous deletion, named *Df(16)1*, encompassing around 1 megabase of the homologous region in mouse causes cardiovascular abnormalities characteristic of the human disease². Here we have used a combination of chromosome engineering and P1 artificial chromosome transgenesis to localize the haploinsufficient gene in the region, *Tbx1*. We show that *Tbx1*, a member of the T-box transcription factor family, is required for normal development of the pharyngeal arch arteries in a gene dosage-dependent manner. Deletion of one copy of *Tbx1* affects the development of the fourth pharyngeal arch arteries, whereas homozygous mutation severely disrupts the pharyngeal arch artery system. Our data show that haploinsufficiency of *Tbx1* is sufficient to generate at least one important component of the DiGeorge syndrome phenotype in mice, and demonstrate the suitability of the mouse for the genetic dissection of microdeletion syndromes.

The *Df1* deletion, hitherto referred to as *Df(16)1*, includes at least 15 genes, several of which might be considered potential candidate genes. The human deletion map is not helpful in selecting candidate genes because there are patients with non-overlapping deletions within the *de22q11* region that are associated with clinically similar phenotypes^{1,3–5}. To map the haploinsufficient gene responsible for the cardiovascular phenotype, we produced nested deletion and duplication alleles using two chromosome engineering strategies^{6,7}. We inserted chromosome engineering cassettes sequentially by homologous recombination into the *Es2el* and *Cdcrel1* genes, located around 700 kilobases (kb) apart (Fig. 1a). Targeting of *Es2el* has been described². A targeting vector for gene *Cdcrel1* was designed to replace exons 2–4, which encode

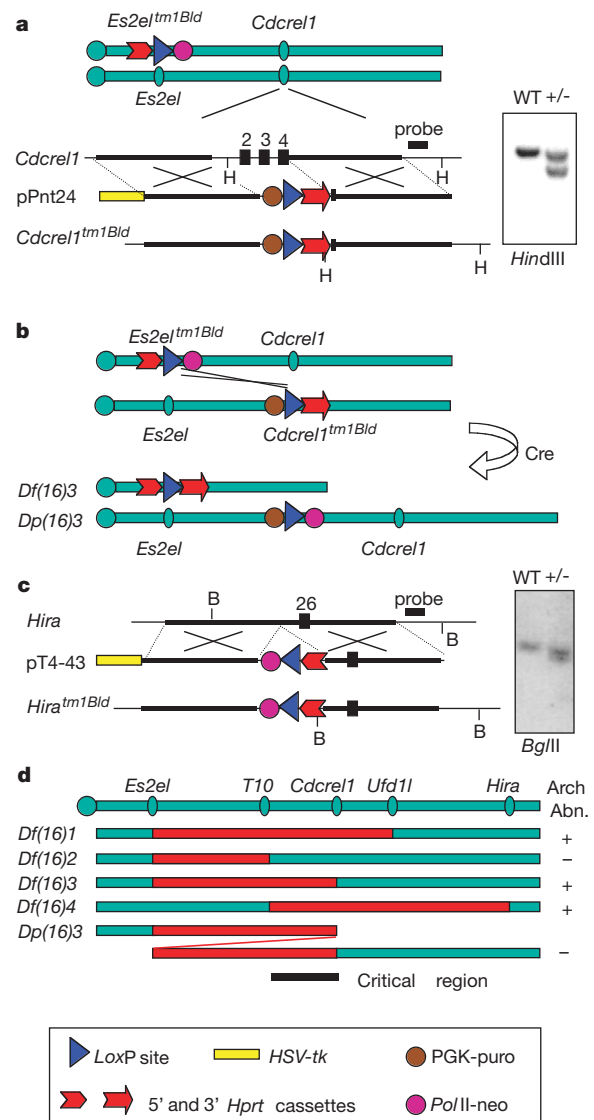


Figure 1 Gene targeting experiments for chromosome engineering. **a**, Targeting of *Cdcrel1*, by homologous recombination, to generate *Cdcrel1^{tm1Bld}*. The *Es2el*-targeted ES cell line used has been described². **b**, Cre-mediated trans-recombination in ES cells to obtain alleles deleted and duplicated for the segment *Es2el–Cdcrel1*. **c**, Targeting of gene *Hira* by homologous recombination to obtain *Hiram1Bld*. Exon 26 is the last exon of *Hira* and contains the polyadenylation signal. *Hira*-targeted ES cells were then used as a substrate for retrovirus-mediated *loxP* insertion. **d**, Diagram of the deletion (single red box) and duplication (double red box) alleles established in mouse line. See text for the full nomenclature of these alleles. On the right, + and – indicate the presence and absence of the fourth PAA abnormalities in embryos carrying this deletion, respectively. WT, wild-type allele; Tar, targeted allele; H, *HindIII*; B, *BglII*; Arch Abn., Abnormalities of the fourth PAAs.

the GTP binding domain, with the *Hprt3'* chromosome engineering cassette in the appropriate orientation (Fig. 1a). Embryonic stem (ES) cells that were double-targeted *in trans* were used to obtain *Cdcrel1* gene knockout mice. These cells were subsequently subjected to Cre-mediated recombination. This generated ES cell lines with reciprocal deletion and duplication chromosomes for the interval *Es2el–Cdcrel1*. The deletion and duplication are termed *del(16)(Es2el–Cdcrel1)* (abbreviated to *Df(16)3*) and *dup(16)(Es2el–Cdcrel1)* (abbreviated to *Dp(16)3*), respectively (Fig. 1b).

We also generated two additional chromosomal deletions using retrovirus-mediated insertion of the *Hprt3'* chromosome engineering cassette⁷. As a substrate for retroviral infection, we used two ES cell lines in which the first cassette (*Hprt5'*) had previously been inserted into the *Es2el* or *Hira* genes (Fig. 1c). The extent of the retroviral-mediated deletions was determined by fluorescence *in situ* hybridization or by cloning the retroviral integration site. Two

engineered chromosomes were established in mice, one with a roughly 500-kb deletion from *Es2el* to *T10*, termed *del(16)(Es2el–T10)* (abbreviated to *Df(16)2*), and one with a roughly 1-megabase (Mb) deletion from *T10* to *Hira*, termed *del(16)(T10–Hira)* (abbreviated to *Df(16)4*; Fig. 1d).

The primary embryological defect underlying most of the cardiovascular abnormalities observed in *Df(16)1/+* mice is defective development of the fourth pharyngeal arch arteries (PAAs)². These defects are 100% penetrant at embryonic day (E) 10.5 (that is, all mutant embryos have at least one abnormal fourth PAA) and can be visualized by intracardiac India ink injection and scored as abnormally small or 'absent' (non-patent to ink). We used ink injection to analyse the PAA phenotype in E10.5 embryos carrying the deletion chromosomes. The characteristic fourth PAA abnormalities were detected in all *Df(16)3/+* and *Df(16)4/+* embryos ($n = 16$ and 9 , respectively). In contrast, no abnormalities were detected in *Df(16)2/+* embryos ($n = 25$), consistent with the lack of

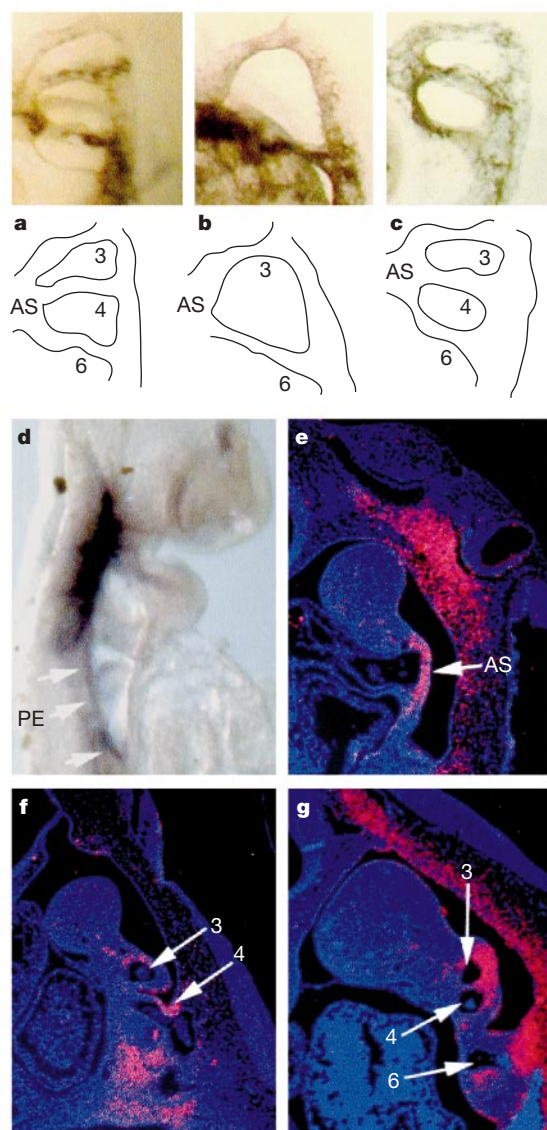


Figure 2 Transgenic rescue of the *Df(16)1/+* phenotype and developmental expression of the *Tbx1* gene in wild-type embryos. **a**, PAAs in an E10.5 wild-type embryo visualized by intracardiac ink injection. **b**, Absence of a fourth PAA in a *Df(16)1/+* embryo at E10.5. **c**, Rescued fourth PAA phenotype in a *Df(16)1/+;Tg+* embryo. AS, Aortic sac; 3, 4 and 6 indicate the third, fourth and sixth PAAs, respectively. **d–g**, RNA *in situ* hybridization using a *Tbx1* riboprobe (antisense) on E9.5 (**d**), E10.25 (**e, f**) and E10.5 (**g**) wild-type embryos. Hybridization with a control (sense) riboprobe produced no specific signal (not shown).

d, Whole-mount *in situ* hybridization. Arrowheads indicate expression in the pharyngeal endoderm (PE). **e**, Parasagittal section of an E10.25 embryo hybridized with a ³⁵S-labelled *Tbx1* riboprobe. Note strong expression in the dorsal wall of the aortic sac (AS). **f**, A more lateral section of the same embryo showing strong expression on the wall of the fourth PAA. **g**, Parasagittal section of an E10.5 embryo showing expression of *Tbx1* in the mesenchyme adjacent to the third, fourth and sixth PAAs.

cardiovascular defects in mice with a similar deletion⁸. These results indicate that the deletion of a segment between genes *T10* and *Cdcrel1* (around 200 kb) is necessary to generate the fourth PAA abnormalities (critical region, Fig. 1d). This segment may include the haploinsufficient gene, but it is possible that the deletion of this segment causes phenotypic abnormalities by deleting regulatory elements of genes distal to *Cdcrel1*. To exclude this, we generated *Df(16)1/Dp(16)3* mice. *Df(16)1/Dp(16)3* E10.5 embryos have normal fourth PAA anatomy ($n = 12$) indicating that the haploinsufficient gene is proximal to *Cdcrel1*, as *Dp(16)3* can compensate only for haploidy of genes located in the interval *Es2el*–*Cdcrel1*.

To localize more precisely the haploinsufficient gene, mouse genomic clones from the *T10*–*Cdcrel1* interval were used to rescue the *Df(16)1/+* phenotype. Transgenic mice carrying one or two copies of P1 artificial chromosome (PAC) 344g21 were bred with *Df(16)1/+* mice and the progeny analysed at E10.5 by intracardiac ink injection. This 140-kb PAC fully rescued the fourth PAA abnormalities and all *Df(16)1/+*;Tg+ embryos were normal ($n = 16$) (Fig. 2a–c). These data indicate that the haploinsufficient gene causing fourth PAA abnormalities and its essential *cis* regulatory elements are located within this transgene. Sequence data analysis of PAC344g21 indicated that the clone includes four genes: *Gnb1l*, *Tbx1*, *Gplbβ* and *Cdcrel1*. The last is not implicated in the haploinsufficiency phenotype because *Cdcrel1*^{−/−} mice, obtained as described above, are healthy and fertile and show no cardiovascular defects. *Gplbβ* (glycoprotein Ib, platelet, β-polypeptide gene) is required for normal platelet function. Although no mouse mutant has been reported for this gene, homozygous loss of function of this locus in humans causes Bernard–Soulier syndrome (OMIM no. 231200, ref. 9), a recessive bleeding disorder that is not associated with congenital heart disease or other *del22q11*/DiGeorge syndrome (DGS)-related clinical findings.

We examined the expression profiles of *Gnb1l* (ref. 10) and *Tbx1* (ref. 11) to investigate whether either was expressed in the cardio/cranial neural crest, heart or pharyngeal arches or pouch endoderm. *Gnb1l* is expressed at high levels in the sclerotome of E9.5 embryos, but at low or undetectable levels in the neuroepithelium and pharyngeal arches at E9.5 and E10.5 (data not shown). Conversely, expression analysis of *Tbx1* corroborated previously reported data, which showed that the gene is expressed in the pharyngeal arches and pouches at E9.5–E12.5 (ref. 11). In particular, at E9.5 *Tbx1* is expressed in the mesenchyme ventral to the hindbrain, the pharyngeal endoderm, pharyngeal arches, sclerotome and, to a lesser extent, in the brain (Fig. 2d). At E10.25, *Tbx1* is strongly expressed in the dorsal wall of the aortic sac and in pharyngeal endoderm and mesenchyme (Fig. 2e). It is also strongly expressed in the third and fourth pharyngeal pouches, and in the medial wall of the forming fourth PAAs (Fig. 2f). At E10.5, expression is clearly evident in the pharyngeal mesenchyme adjacent to the third, fourth and sixth PAAs (Fig. 2g).

To establish whether heterozygous deletion of *Tbx1* alone is sufficient to cause fourth PAA abnormalities, we generated a mutant allele, *Tbx1*^{tm1Bld}, by inserting a *lacZ* reporter gene into exon 5, which encodes part of the T-box domain (Fig. 3a). Embryos heterozygous for this allele (abbreviated as *Tbx1*^{+/-}) were collected at E10.5 and the PAAs were analysed by ink injection. *Tbx1*^{+/-} embryos exhibited fourth PAA abnormalities identical to those of *Df(16)1/+* embryos ($n = 14$) and with the same, full penetrance (Fig. 3c). *Tbx1*^{−/−} mutants revealed severe disruption of the PAA system (Fig. 3d). In these embryos, the third, fourth and sixth PAAs were absent; instead, the dorsal aortae connected directly with the aortic sac (Fig. 3d), suggesting a severe developmental impairment of the pharyngeal arches. Compound heterozygous, *Df(16)1/Tbx1*^{tm1Bld} embryos exhibited an identical phenotype

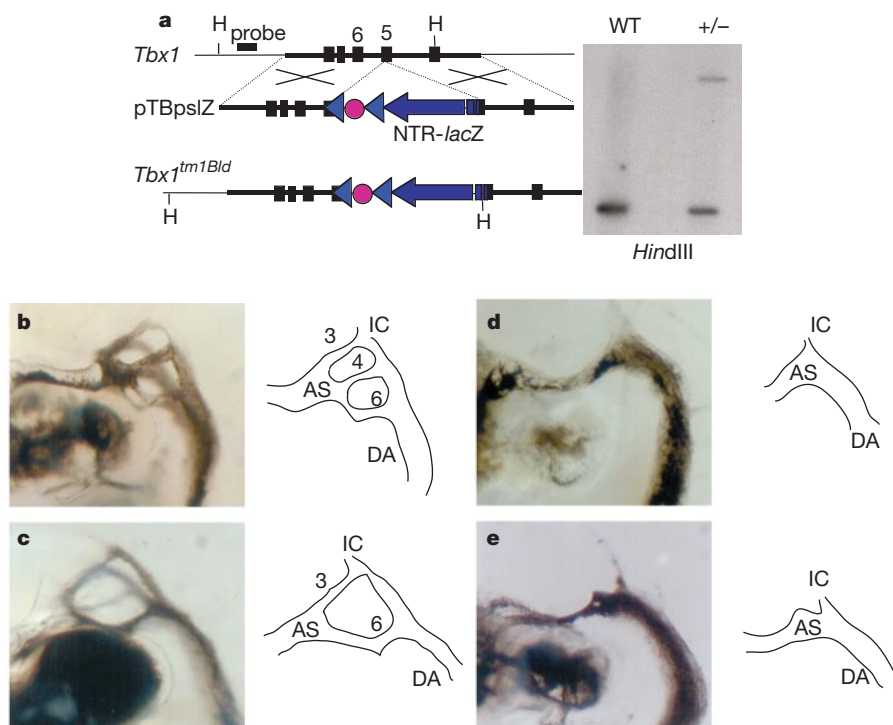


Figure 3 *Tbx1* gene mutagenesis and resulting pharyngeal arch artery phenotype. **a**, *Tbx1* gene targeting and generation of the *Tbx1*^{tm1Bld} allele. Exons are named according to published data¹². **b**, Typical PAA morphology in a wild-type E10.5 embryo as visualized by intracardiac ink injection. **c**, Absence of the left fourth PAA in a *Tbx1*^{+/-} embryo at E10.5. The fourth PAA on the contralateral side appears normal. **d**, PAA morphology in a *Tbx1*^{−/−}

embryo; none of the PAAs are recognizable and the dorsal aorta connects directly with the aortic sac. **e**, Compound heterozygous *Df(16)1/Tbx1*^{tm1Bld} embryo; the pattern is essentially identical to that of *Tbx1*^{−/−} embryos. AS, aortic sac; DA, dorsal aorta; IC, internal carotid artery.

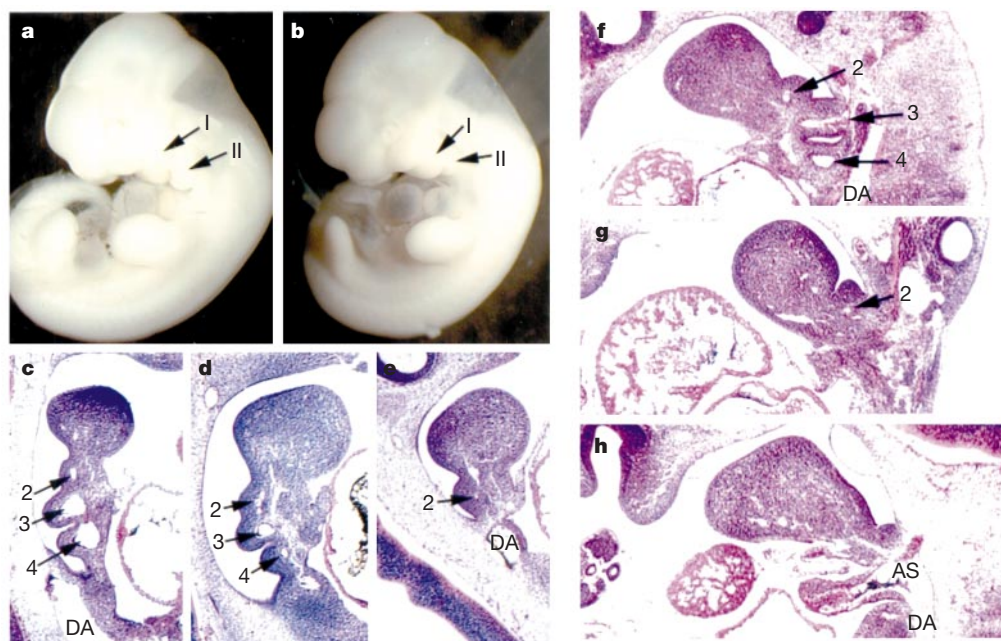


Figure 4 The pharyngeal arch phenotype in *Tbx1* mutant and compound heterozygous embryos. **a**, External appearance of wild-type, and **(b)** *Tbx1*^{-/-} E11 embryos. Note the hypoplasia of the second pharyngeal arch (II) in the mutant embryo. **c–e**, Parasagittal histological sections through the pharyngeal arch region in wild-type **(c)**, *Tbx1*^{+/-} **(d)**, and *Df(16)1/Tbx1*^{tm1Bld} **(e)** embryos at E10.5. Note the hypoplasia of the fourth PAA in the *Tbx1*^{+/-} embryo, and the reduced size of the second PAA and absence of the third and

fourth pharyngeal arches and arch arteries in the *Df(16)1/Tbx1*^{tm1Bld} embryo. **f–h**, Parasagittal sections through the pharyngeal arch region in wild-type **(f)** and *Tbx1*^{-/-} **(g, h)** embryos at E11. Note that in common with *Df(16)1/Tbx1*^{tm1Bld} embryos, *Tbx1*^{-/-} embryos lack pharyngeal arch and arch artery development distal to the second arch **(g)** and that the dorsal aorta connects directly to the aortic sac. DA, dorsal aorta; AS, aortic sac.

(Fig. 3d). *Tbx1*^{-/-} and *Df(16)1/Tbx1*^{tm1Bld} embryos were viable and of similar size to their wild-type littermates. However, the second pharyngeal arches were small by external inspection (Fig. 4a, b).

Histological sections confirmed the dosage-dependent effects of *Tbx1* mutation (Fig. 4c–e), demonstrating hypoplasia of the fourth PAAs in *Tbx1*^{+/-} mutants (Fig. 4d), which is an identical phenotype to *Df1/+* mutants (not shown), and the disruption of the third, fourth and sixth arch and arch arteries in compound heterozygous *Df(16)1/Tbx1*^{tm1Bld} and in *Tbx1*^{-/-} animals (Fig. 4e–h). As predicted by the ink injection experiments, the dorsal aortae originate directly from the aortic sac in *Tbx1*^{-/-} (Fig. 4h) and *Df(16)1/Tbx1*^{tm1Bld} (not shown) animals. The pharyngeal endoderm was morphologically normal. We could not detect any phenotypic difference between *Tbx1*^{-/-} and *Df(16)1/Tbx1*^{tm1Bld} mutants, suggesting that haploinsufficiency of genes other than *Tbx1* from the *Df(16)1* segment is unlikely to contribute to the PAA phenotype.

Thus *Tbx1* is a dosage-sensitive gene, which when haploid results in fourth PAA abnormalities. Loss of function of *Tbx1*, as tested in *Tbx1*^{-/-} or *Df(16)1/Tbx1*^{tm1Bld} embryos, leads to disruption of PAAs 3–6, possibly secondary to lack of segmentation of the distal pharyngeal arches and pouches. Great artery defects involving third and sixth PAA derivatives are almost never seen in *del22q11* DGS patients, suggesting that our haploinsufficiency models reflect more closely the aortic arch defects than do the homozygous mutations. *TBX1* is deleted in most DGS patients¹², but whether haploinsufficiency of this gene alone causes the entire phenotype associated with *del22q11*/DGS remains to be determined. However, as the homozygous mouse mutant presents with developmental defects of the second pharyngeal arch and of the third and fourth pharyngeal pouches, it is possible that in humans *TBX1* haploinsufficiency causes the craniofacial, thymic and parathyroid anomalies seen in patients.

Mutational analysis of the coding region of the *TBX1* gene in over 100 patients with a clinical suspicion of *del22q11*/DGS but without

deletion or other detectable rearrangements has so far failed to identify *Tbx1* gene mutations (P.J.S., unpublished data). Negative results were reported for a smaller cohort of patients¹². Non-overlapping deletions within the *del22q11* region in patients with apparently similar phenotypes suggest the presence of long-range *cis*-acting genetic elements. However, our data do not support this hypothesis because a region of 140 kb contains all of the genetic elements necessary to rescue the aortic arch phenotype in mice. However, it is possible that these hypothetical long-range *cis*-acting elements are not conserved in mice. Alternatively, non-overlapping deletions in humans may affect distinct loci, the deletion of which may independently cause a *del22q11*/DGS-like phenotype, perhaps through different mechanisms.

Tbx1 is the first dosage-sensitive gene identified in the *del22q11*/DGS region and is the fourth member of the T-box family of transcription factor genes shown to be haploinsufficient in mammals, in addition to *brachyury*, *TBX3*, and *TBX5* (refs 13–16). We hypothesize that *Tbx1* is involved in pharyngeal endoderm–mesenchyme interactions that lead to segmentation of the distal pharyngeal arches. It is unlikely that the phenotype observed in our mutants is due to a primary defect of neural crest cells, because these cells are not required for segmentation of the pharyngeal pouches^{17,18}. The identification of genes targeted by *Tbx1* should provide insight into pharyngeal arch development in mammals and into the pathogenesis of the many birth defects that derive from abnormal development of these structures. □

Methods

Gene targeting and transgenic mice

All targeting experiments were performed using the ES cell line AB2.2 derived from *Hprt*-deficient 129SvEvBrd mice¹⁹. *Cdcrel1* targeting was performed on an *Es2el*-targeted cell line as described². The chromosome engineering cassettes used for *Es2el* and *Hira* targeting (*Hprt5'*), and for *Cdcrel1* targeting (*Hprt3'*), have been described⁶. *Tbx1* gene targeting was performed using a *lacZ* reporter gene (coding a nuclear localization signal) preceded

by an internal ribosome entry site, and followed by an SV40 polyadenylation signal, and a floxed PGK-*neo* cassette (a gift from R. Behringer). The retroviral-mediated insertion of the *Hprt3'* chromosome engineering cassette was performed as described⁷. Deleted clones were analysed by fluorescence *in situ* hybridization²⁰ to determine the extent of the deletion, using bacterial artificial chromosome (BAC) clones previously mapped to the murine locus homologous to *del22q11* (refs 21–25). For deletion *Df(16)4*, the deletion breakpoint was also mapped by cloning and sequencing the retroviral insertion point using inverse PCR⁷. To obtain transgenic mice, DNA from PAC344g21 (library RPCI21 made by K. Osoegawa and P. de Jong, generated from 129/SvEvTACfBr mouse DNA, and obtained from HGMP) was linearized with lambda terminase, purified, and injected into pronuclei of one-cell mouse embryos, which were then transferred into the oviducts of foster mothers. The presence of the transgene was detected at weaning using PCR primer pairs specific for the PAC vector-insert junctions or specific for the vector only. The founder used for breeding carried one or two copies of the transgene as detected by semi-quantitative Southern blotting analysis. All of the targeted heterozygous mutants and transgenic mice were viable.

To determine the gene content of the PAC clone 344g21, end sequences were determined and mapped to sequence files recorded in GenBank (AC003066 and AC008019). The PAC sequence was then compared to the sequences recorded in public databases, including expressed sequence tags, using publicly available BLAST programs. The sequence of a *Tbx1* complementary DNA is deposited in GenBank under the accession number AF326960.

Mouse strains, breeding and genotyping

All mutant mouse lines were examined in a mixed genetic background 129SvEvBrd (129S5) × C57BL/6, with the exception of the transgenic line Tg(344g21)1Bld, which was obtained in a mixed C57BL/6 × SJL genetic background, then back-crossed twice with C57BL/6 mice before performing rescue experiments. This was necessary as the penetrance of the haploinsufficiency phenotype is affected by genetic background (our unpublished data). Two rounds of back-crossing were sufficient to restore the full penetrance of the fourth PAA phenotype at E10.5.

Embryos and mice were genotyped using DNA extracted from yolk sacs or tail biopsies, respectively. Analyses were performed by PCR using the following primer pairs: *Df(16)1* deletion: ufdtar2-f: 5'-TCITTTGTCAGCAGTTCCTTT-3', ufdtar2-r: 5'-TGGGCAATT GTTAAATCTCC-3'. *Df(16)2*: es2tar-f: 5'-GCCTTTGTTTCAAGCTCCA-3', es2tar-r: 5'-TACCGGTGGATGTGGAATGT-3'. *Dp(16)3*: es2dup3: 5'-GGGAGGATTGGGAAG ACAAT-3', es2dup4: 5'-GGGCAGAGGAAGAGCCTACT-3'. *Df(16)3*: cdc24tar-f: 5'-AACTTGACCCCTTCTCCTC-3', cdc24tar-r: 5'-GTGGCTCTATGGCTTCTGA-3'. *Df(16)4*: tupwt-f: 5'-GGGCAGGAGGACTGCATATT-3', tuptar-r: 5'-TACCGGTGGAT GTGGAATGT-3'. *Tbx1*^{1m1Bld}: tbxtar2-f: 5'-CAGAGAAGGGTCGCCTACAT-3', tbxtar2-r: 5'-TCGACTAGAGCTTGCAGAAC-3'. Tg(344g21)1Bld: pacsp6-f: 5'-TAGTCAATTCGG GAGGATCG-3', pacsp6-r: 5'-ATGAGAGACGCCGTACGTTT-3'.

Ink injection, histology and RNA *in situ* hybridization

For India ink injection, embryos were collected and injected intracardially with India ink, and fixed in PBS-buffered 4% formaldehyde. Embryos were then dehydrated and cleared in 1:1 benzyl benzoate:methyl salicylate.

For histology and radioactive RNA *in situ* hybridization, embryos were fixed in PBS containing 4% paraformaldehyde, embedded in paraffin and cut into 10-µm sections. Section and whole-mount *in situ* hybridization experiments were performed according to published protocols²⁶. We used a previously described probe to detect *Tbx1* expression¹¹. Histological stainings were performed using haematoxylin and eosin.

Received 12 December 2000; accepted 7 February 2001.

- Scambler, P. J. The 22q11 deletion syndromes. *Hum. Mol. Genet.* **9**, 2421–2426 (2000).
- Lindsay, E. A. *et al.* Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* **401**, 379–383 (1999).
- O'Donnell, H., McKeown, C., Gould, C., Morrow, B. & Scambler, P. Detection of an atypical 22q11 deletion that has no overlap with the DiGeorge syndrome critical region. *Am. J. Hum. Genet.* **60**, 1544–1548 (1997).
- Amati, F. *et al.* Atypical deletions suggest five 22q11.2 critical regions related to the DiGeorge/velocardio-facial syndrome. *Eur. J. Hum. Genet.* **7**, 903–909 (1999).
- McQuade, L. *et al.* Patient with a 22q11.2 deletion with no overlap of the minimal DiGeorge syndrome critical region (MDGCR). *Am. J. Med. Genet.* **86**, 27–33 (1999).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Su, H., Wang, X. & Bradley, A. Nested chromosomal deletions induced with retroviral vectors in mice. *Nature Genet.* **24**, 92–95 (2000).
- Puech, A. *et al.* Normal cardiovascular development in mice deficient for 16 genes in 550 kb of the velocardiofacial/DiGeorge syndrome region. *Proc. Natl Acad. Sci. USA* **97**, 10090–10095 (2000).
- McKusick, V. *On Line Mendelian Inheritance in Man* (<http://www3.ncbi.nlm.nih.gov/Omim/>) (2000).
- Gong, L., Liu, M., Jen, J. & Yeh, E. T. GNB1L, a gene deleted in the critical region for DiGeorge syndrome on 22q11, encodes a G-protein beta-subunit-like polypeptide(1). *Biochim. Biophys. Acta* **1494**, 185–188 (2000).
- Chapman, D. L. *et al.* Expression of the T-box family genes, *Tbx1*–*Tbx5*, during early mouse development. *Dev. Dyn.* **206**, 379–390 (1996).
- Chieffo, C. *et al.* Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse *Tbx1* gene. *Genomics* **43**, 267–277 (1997).
- Smith, J. Brachyury and the T-box genes. *Curr. Opin. Genet. Dev.* **7**, 474–480 (1997).
- Bamshad, M. *et al.* Mutations in human *TBX3* alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16**, 311–315 (1997).
- Basson, C. T. *et al.* Mutations in human *TBX5* cause limb and cardiac malformation in Holt-Oram syndrome. *Nature Genet.* **15**, 30–35 (1997).
- Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in *TBX5*, a member of the Brachyury (T) gene family. *Nature Genet.* **15**, 21–29 (1997).

- Kirby, M. L., Hunt, P., Wallis, K. & Thorogood, P. Abnormal patterning of the aortic arch arteries does not evoke cardiac malformations. *Dev. Dyn.* **208**, 34–47 (1997).
- Begbie, J., Brunet, J. F., Rubenstein, J. L. & Graham, A. Induction of the epibranchial placodes. *Development* **126**, 895–902 (1999).
- Matzuk, M. M., Finegold, M. J., Su, J. G., Hsueh, A. J. & Bradley, A. Alpha-inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature* **360**, 313–319 (1992).
- Baldini, A. & Lindsay, E. A. Mapping human YAC clones by fluorescence *in situ* hybridization using Alu-PCR from single yeast colonies. *Methods Mol. Biol.* **33**, 75–84 (1994).
- Botta, A., Lindsay, E. A., Jurecic, V. & Baldini, A. Comparative mapping of the DiGeorge syndrome region in mouse shows inconsistent gene order and differential degree of gene conservation. *Mamm. Genome* **8**, 890–895 (1997).
- Galili, N. *et al.* A region of mouse chromosome 16 is syntenic to the DiGeorge, velocardiofacial syndrome minimal critical region. *Genome Res.* **7**, 17–26 (1997).
- Puech, A. *et al.* Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mouse reveals complex changes in gene organization. *Proc. Natl Acad. Sci. USA* **94**, 14608–14613 (1997).
- Sutherland, H. F., Kim, U. J. & Scambler, P. J. Cloning and comparative mapping of the DiGeorge syndrome critical region in the mouse. *Genomics* **52**, 37–43 (1998).
- Lund, J. *et al.* Sequence-ready physical map of the mouse chromosome 16 region with conserved synteny to the human velocardiofacial syndrome region on 22q11.2. *Mamm. Genome* **10**, 438–443 (1999).
- Albrecht, U., Eichele, G., Helms, J. A. & Lu, H. C. in *Molecular and Cellular Methods in Developmental Toxicology* (ed. Daston, G. P.) 23–48 (CRC, New York, 1997).

Acknowledgements

We thank R. Behringer and P. Hastings for discussions and reading of the manuscript, and Y.-C. Cheah, S. Carattini-Rivera, G. C. Schuster, S.-S. Cheah and C. Roberts for technical contributions. A.B. is an investigator with the Howard Hughes Medical Institute. This work was funded in part by grants from the National Heart, Lung and Blood Institute and (to A.B.), from the American Heart Association Texas Affiliate (to E.A.L.), and from the British Heart Foundation and Wellcome Trust (to P.J.S.). Support from the Baylor Mental Retardation Research Center is also acknowledged.

Correspondence and requests for materials should be addressed to A.B. (e-mail: baldini@bcm.tmc.edu).

Visualizing the generation of memory CD4 T cells in the whole body

R. Lee Reinhardt*, Alexander Khoruts†, Rebecca Merica*, Traci Zell* & Marc K. Jenkins*

* Department of Microbiology, and † Department of Medicine, Center for Immunology, University of Minnesota Medical School, MMC 334, 420 Delaware Street S.E., Minneapolis, Minnesota 55455, USA

It is thought that immunity depends on naïve CD4 T cells that proliferate in response to microbial antigens^{1–4}, differentiate into memory cells that produce anti-microbial lymphokines^{5,6}, and migrate to sites of infection^{7,8}. Here we use immunohistology to enumerate individual naïve CD4 T cells, specific for a model antigen, in the whole bodies of adult mice. The cells resided exclusively in secondary lymphoid tissues, such as the spleen and lymph nodes, in mice that were not exposed to antigen. After injection of antigen alone into the blood, the T cells proliferated, migrated to the lungs, liver, gut and salivary glands, and then disappeared from these organs. If antigen was injected with the microbial product lipopolysaccharide, proliferation and migration were enhanced, and two populations of memory cells survived for months: one in the lymph nodes that produced the growth factor interleukin-2, and a larger one in the non-lymphoid tissues that produced the anti-microbial lymphokine interferon-γ. These results show that antigen recognition in the context of infection generates memory cells that are specialized to proliferate in the secondary lymphoid tissues or to fight infection at the site of microbial entry.

It has been difficult to track the behaviour of naïve CD4 T cells that are specific for a single antigen because these cells are extremely

rare before immunization, and because of a lack of methods that allow accurate sampling of T cells in all parts of the body⁹. We solved the frequency problem here by injecting normal mice with several million naïve CD4 T cells from a T-cell antigen receptor (TCR) transgenic mouse line, specific for a known antigen. The sampling problem was solved by tracking the injected T cells in thin sections through the whole bodies of recipient mice. CD4 T cells

from the OT-II transgenic C57BL/6.PL mouse line¹⁰ specific for a peptide from chicken ovalbumin were injected intravenously into histocompatible adult C57BL/6 mice. The OT-II cells were identified on the basis of expression of the Thy-1.1 molecule, which differs from the Thy-1.2 molecule expressed by C57BL/6 mice. The recipient mice were given an intravenous (i.v.) injection with ovalbumin peptide 323–339 (referred to hereafter as ovalbumin

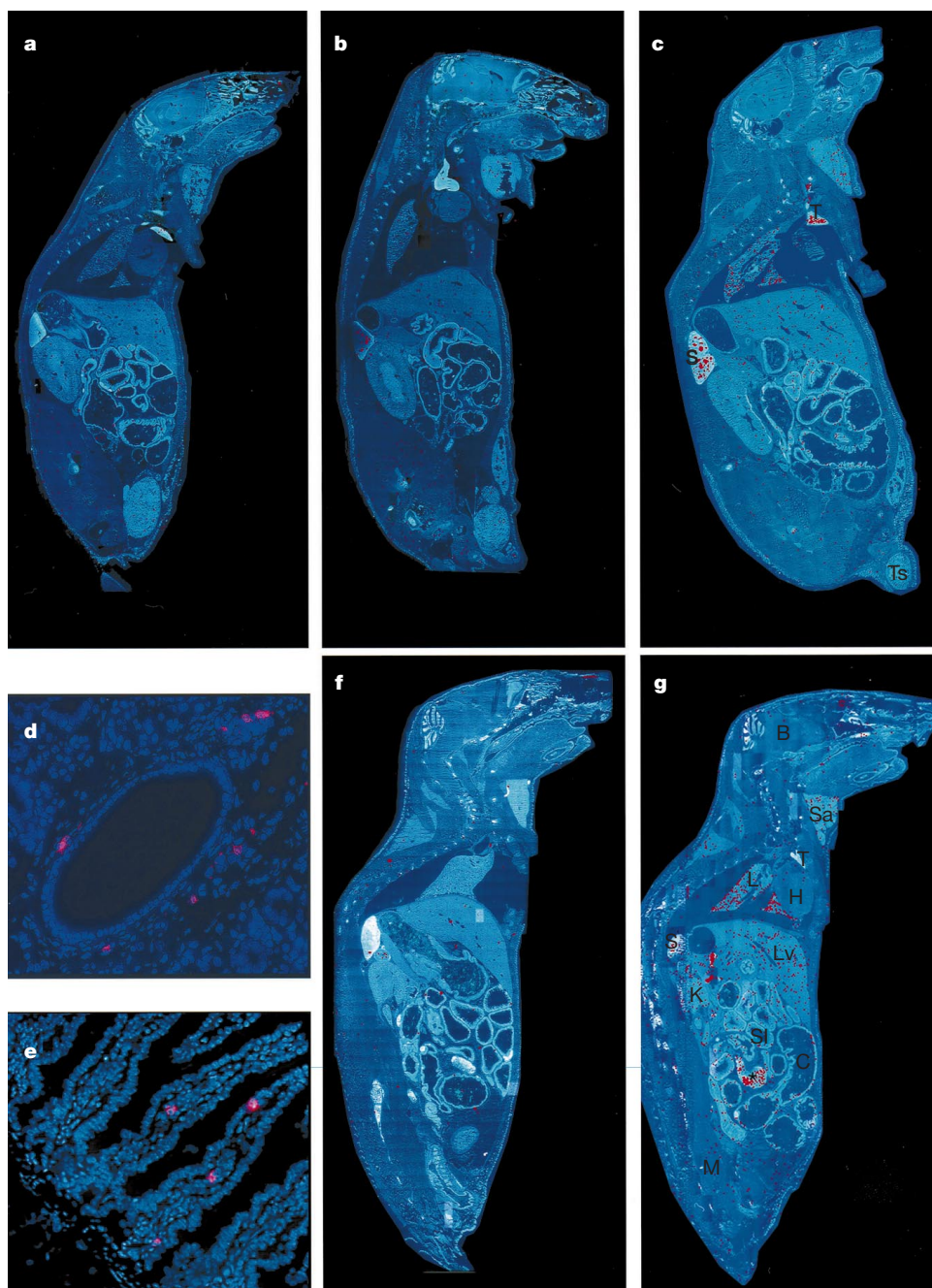


Figure 1 Detection of antigen-specific T cells in whole-body sections. Images from anti-Thy-1.1/Cy3-stained sections through the following mice are shown: **a**, Normal C57BL/6 mouse that did not receive OT-II cells; **b**, C57BL/6 mouse that received 3×10^6 OT-II cells intravenously 4 days before sacrifice; and **c**, C57BL/6 mouse that received OT-II cells, given an i.v. injection the next day with ovalbumin peptide and LPS, and killed 3 days later. **d**, **e**, Images showing $\times 100$ views of the salivary gland (**d**) and lamina propria (**e**) of the small intestine, respectively, from OT-II recipients that were injected with ovalbumin peptide and LPS. **f**, **g**, Composite images from a normal BALB/c mouse (**f**), and a BALB/c bone marrow chimera (**g**), in which roughly 1% of the T cells were derived from a

DO11.10 SCID TCR transgenic mouse, and given an i.v. injection with 100 μ g ovalbumin peptide and 50 μ g LPS 5 days before sacrifice. In this case, the red signal represents staining with a Cy3-labelled anti-clonotypic antibody, KJ1-26, that uniquely recognizes the DO11.10 TCR. No DO11.10 T cells were detected in the thymus of **g** because only cortex is present on this section, and ovalbumin peptide/LPS-activated T cells only entered the medulla (**c**). B, brain; H, heart; K, kidney; L, lung; C, colon; Lv, liver; M, muscle; S, spleen; Sa, salivary glands; Si, small intestine; T, thymus; Ts, testes; asterisks, lymph nodes. Images **a–e** are derived from males; **f** and **g** from females.

peptide) alone or with lipopolysaccharide (LPS). We compared these two regimens because administration of antigen under inflammatory conditions, such as those produced by microbes or LPS, induces immunological memory, whereas injection of antigen alone does not^{11–13}.

Mice were killed at various times after injection of ovalbumin peptide and perfused to remove cells from blood vessels. Mice were then frozen in liquid nitrogen and thin sections were cut as described¹⁴. Whole-mouse sections were stained with Cy3-labelled anti-Thy-1.1 monoclonal antibody to identify the OT-II cells, and 4,6-diamidino-2-phenylindole (DAPI), a DNA stain that highlights the anatomic features of individual organs. Sets of digital images covering completely each section in the Cy3 (red) and DAPI (blue) channels were then assembled to create a composite image of each whole-mouse section. The number of OT-II cells present in individual organs was calculated on the basis of the difference between the number of red pixels present in a tissue on an anti-Thy-1.1-stained section from a recipient of OT-II cells, and the background value for that tissue from a mouse that did not receive OT-II cells.

Some artefactual red objects (~170) were detected on anti-Thy-1.1-stained sections through C57BL/6 mice that did not receive OT-II cells (Fig. 1a). This was a relatively low value given that only one in four of the ~700 images that comprise the composite shown in Fig. 1a contained a single spot of artefactual staining. Secondary lymphoid organs (spleen (Fig. 1b), lymph nodes and Peyer's patches (data not shown)) were the only organs in sections of naïve or LPS-injected recipients of OT-II cells that contained significantly more red objects than the comparable organs from a normal C57BL/6 mouse (Fig. 2). Although a greater than basal level of red objects were detected in the liver and thymus of naïve OT-II recipients, these values were only slightly above the limit of detection (Fig. 2). Therefore, naïve OT-II cells resided primarily in the secondary lymphoid organs, and LPS alone did not affect this residence pattern.

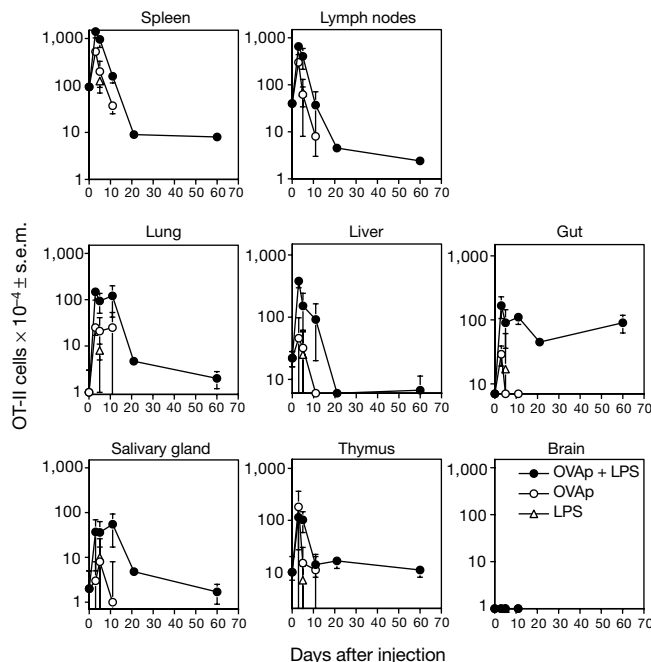


Figure 2 Tissue distribution of antigen-specific T cells during a primary immune response. Each panel shows the mean number of OT-II cells $\pm$ s.e.m. ($n = 2-5$ mice from five independent experiments) in the indicated tissues of mice given an i.v. injection with ovalbumin peptide (OVAp) alone (open circles), ovalbumin peptide and LPS (closed circles), or LPS alone (open triangles). The minimum value for the y-axis on each graph is the limit of detection for that tissue.

As expected from past results¹⁵, the number of OT-II cells increased in the spleen and lymph nodes 3 days after injection of ovalbumin peptide alone, and to a greater extent after injection of ovalbumin peptide plus LPS (Fig. 2). Most of the cells then disappeared from the secondary lymphoid organs by day 11 in both cases (Figs 2 and 3a). By day 3 after injection, OT-II cells appeared in the interstitial spaces of the liver, lung, salivary gland (Fig. 1d), lamina propria of the gut (Fig. 1e) and the medulla of the thymus (Fig. 1c). More OT-II cells appeared in most of these tissues in mice injected with ovalbumin peptide plus LPS than in mice injected with ovalbumin peptide alone (Fig. 2). A similar appearance of antigen-specific CD4 T cells in tissues other than secondary lymphoid tissues was observed 5 days after injection of ovalbumin peptide plus LPS in a different system involving the DO11.10 ovalbumin peptide-specific TCR transgenic line¹⁶ (compare Fig. 1f and g).

The OT-II cells disappeared completely from the non-lymphoid tissues and the thymus between 3 and 11 days in mice injected with ovalbumin peptide alone (Figs 2 and 3b). The number of OT-II cells present in these tissues also fell markedly over this time in mice injected with ovalbumin peptide plus LPS (Figs 2 and 3b). The number of OT-II cells, however, stabilized 21 days after this injection, such that the same number of cells were present in the lungs, salivary gland, gut and thymus on day 60 (Fig. 3b). The gut experienced the smallest loss of OT-II cells between days 11 and 21, and was the largest reservoir of OT-II cells after day 21 (Fig. 2). Ovalbumin-specific CD4 T cells did not appear at greater than background levels in the brain (Figs 1c, g and 2), muscle, heart or testes (Fig. 1c, g), at any time after injection of ovalbumin peptide alone or with LPS. Therefore, naïve CD4 T cells resided primarily in secondary lymphoid tissues, proliferated there in the presence of antigen, and then migrated into other tissues. The presence of a microbial product enhanced antigen-driven T-cell proliferation in the secondary lymphoid tissues and the persistence of antigen-specific T cells in other tissues.

All of the ovalbumin-specific CD4 T cells that were present in the lungs, spleen and lymph nodes 11 days after immunization with ovalbumin peptide plus LPS had the phenotype of memory cells⁶. These cells had divided in the past (see Table 1) and expressed low amounts of CD45RB⁹, but lacked markers of acute activation such as CD25 (Table 1), and were not producing interleukin-2 (IL-2) or

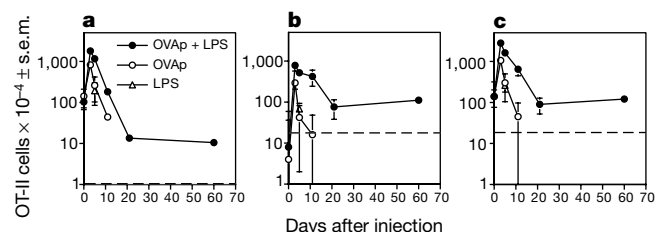


Figure 3 Quantification of antigen-specific T cells. **a–c**, Mean number of OT-II cells $\pm$ s.e.m. ($n = 2-5$ mice from five independent experiments) in the secondary lymphoid tissues (sum of spleen and lymph nodes) (**a**), other tissues (sum of the brain, salivary glands, kidney, thymus, liver, lung and gut) (**b**), and the whole body (secondary lymphoid plus other tissues) (**c**), of mice injected with ovalbumin peptide alone (open circles), ovalbumin peptide and LPS (closed circles), or LPS alone (open triangles). The number of OT-II cells in the non-lymphoid tissues in ovalbumin peptide/LPS-injected mice is probably an underestimate because skin and bone marrow could not be analysed due to the small fraction of their total volume present on any single section. The limits of detection are indicated by a dotted line. These were calculated by summing the limits of detection for the individual tissues in question. It was possible to achieve values lower than the summed limit of detection (note the day 0 values in **b**) because in some tissues the difference between the red pixel value from an OT-II recipient and the negative control yielded a negative number.

interferon- γ (IFN- γ) (Fig. 4, legend). The memory CD4 T cells in the lungs 11 (Fig. 4c) or 60 (Fig. 4d) days after immunization were 5–10 times more likely to produce IFN- γ *in vivo* 2 h (the time of peak lymphokine *in vivo* production by ovalbumin-specific memory cells¹⁷) after i.v. injection of ovalbumin peptide than those in the lymph nodes of the same mice. Memory CD4 T cells in the liver were also much more efficient IFN- γ producers than lymph node cells (data not shown). *In vivo* lymphokine production by the memory CD4 T cells in the gut could not be analysed because of the harsh and lengthy procedure required to isolate these cells.

The frequency of IFN- γ -producing cells in the spleen was intermediate between the frequencies observed in the lungs and the lymph nodes (Fig. 4c, d). The differences in IFN- γ production were not caused by differences in the amount of ovalbumin peptide deposited in the tissues after i.v. injection, because the ovalbumin-specific memory CD4 T cells in the lymph nodes and spleen were more likely to produce IL-2 than those in the lungs (Fig. 4c, d). This point was further supported by the finding that a greater percentage of lung-resident, ovalbumin-specific memory CD4 T cells produced IFN- γ than their lymph node counterparts when stimulated *in vitro* in a peptide-independent fashion with calcium ionophore and phorbol ester ($21.7 \pm 5.2\%$ versus $8.6 \pm 4.9\%$, respectively; mean $\pm$ s.d.). These results show that the ovalbumin-specific memory CD4 T cells that reside in lymph nodes versus non-lymphoid tissues have different functional capabilities.

Although it is reasonably well accepted that naïve CD4 T cells in adults reside primarily in lymphoid tissues^{7,8,18,19}, this point is still

Table 1 Phenotype of antigen-specific CD4 T cells 11 days after immunization

Immunization	Tissue	Divided (%)	CD25* (%)
Ova/LPS	Lymph node	>99	9.8 $\pm$ 2.0
Ova/LPS	Spleen	>99	6.7 $\pm$ 4.3
Ova/LPS	Lungs	>99	5.6 $\pm$ 3.3
Ova/LPS, challenged	Spleen	>99	71.8 $\pm$ 2.6
None	Lymph node	2.2 $\pm$ 1.3	5.4 $\pm$ 1.8

Recipients of DO11.10 CD4 T cells were immunized with ovalbumin (Ova) plus LPS. Eleven days later, the cell division history and phenotype of DO11.10 T cells in the lungs, lymph nodes and spleen were determined by flow cytometry. Some mice were challenged with an i.v. injection of ovalbumin peptide and analysed 1 day later as a positive control for acute activation of DO11.10 T cells. Unimmunized DO11.10 *Rag2*^{-/-} mice were used to assess the phenotype of naïve DO11.10 T cells. Values represent the percentage of DO11.10 cells with evidence of CFSE loss or high levels of CD25 (mean $\pm$ s.d., $n = 2$ –3 mice from separate experiments).

controversial because of the instability of the surface marker changes used to identify these cells in polyclonal populations^{20,21}. The whole-body analysis reported here clearly supports the contention that naïve CD4 T cells reside mainly in secondary lymphoid tissues and have poor access to other tissues.

Injection of ovalbumin peptide alone caused the specific CD4 T cells to proliferate in the secondary lymphoid tissues and then decline dramatically over the next week (Fig. 3a). This disappearance may be explained in part by emigration, because the T cells appeared in many non-lymphoid tissues during this time frame (Fig. 3b). However, the complete loss of T cells from the non-lymphoid tissues is most probably caused by cell death, because it occurred as the total number of cells in the body declined (Fig. 3c). The nearly complete loss of antigen-specific CD4 T cells from the non-lymphoid tissues, together with the previously described lymphokine production defects in the few cells that survive in lymphoid tissues¹⁷, probably explain why immunity does not result from injection of antigen alone^{12,13}.

The microbial product LPS influenced the antigen-stimulated CD4 T-cell response at several levels. The T cells proliferated to a greater extent in the secondary lymphoid tissues, probably as a result of inflammatory cytokines that enhance T-cell proliferation^{15,22,23}. Furthermore, our results show that exposure to antigen under inflammatory conditions fosters the survival of memory CD4 T cells in the secondary lymphoid tissues, and to an even larger degree, in the non-lymphoid tissues, especially the lamina propria of the gut. The constitutive inflammation created in the gut by commensal microbes may make this a particularly nurturing environment for memory cells. Indeed, it has been shown that the memory CD8 T cells in the gut are maintained in a higher state of activation than memory cells in the lymphoid tissues^{24,25}.

The functional differences that we observed suggest that the cells that survive in the lymph nodes are a different subset of memory cells than those that survive in the non-lymphoid tissues. A similar conclusion has been reached²⁶ where it was found that blood contains memory-phenotype T cells that express lymph-node homing molecules and do not produce IFN- γ , and others that lack lymph-node homing molecules and do produce IFN- γ . These results, together with ours, suggest that one subset of memory cells recirculates primarily through lymph nodes, whereas another recirculates primarily through non-lymphoid tissues, both through the blood. The spleen, which samples the blood, would thus contain both types of memory cells and would explain the intermediate IFN- γ production potential of the memory cells in this organ. The high frequency of IL-2-producing cells in the lymph-node-seeking population suggests that its function is to produce more memory cells. In contrast, the memory T cells that recirculate through the non-lymphoid tissues may participate directly in protective immunity by producing anti-microbial lymphokines at sites of infection. If so, then current understanding of protective immunity, which is derived mainly from studies of lymphocytes from blood or

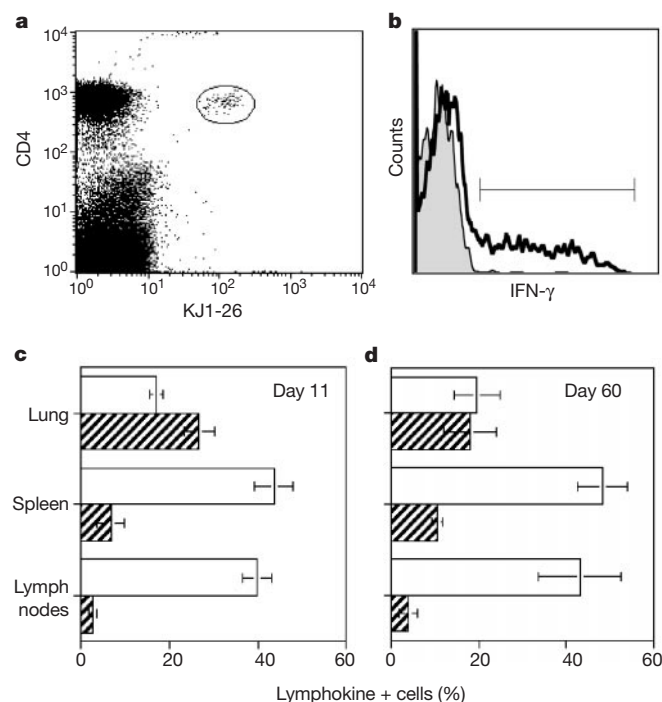


Figure 4 Lymphokine production by antigen-specific CD4 T cells. Recipients of naïve DO11.10 or DO11.10 *Rag2*^{-/-} T cells were injected with ovalbumin peptide plus LPS. Eleven or sixty days later, we used intracellular staining to detect lymphokines in DO11.10 T cells from various tissues after *in vivo* stimulation. **a**, Gate used to identify DO11.10 cells (from the lungs in this example). **b**, IFN- γ staining in DO11.10 cells from the lungs of mice injected 2 h earlier with ovalbumin peptide (open histogram, thick line), or unchallenged mice (shaded histogram, thin line). **c**, **d**, Fraction of DO11.10 T cells containing IFN- γ (striped bars) or IL-2 (open bars) from the indicated tissues of mice injected 2 h earlier with ovalbumin peptide. Values represent the mean percentages ($\pm$ s.d., $n = 4$ –7) of DO11.10 T cells staining for IFN- γ or IL-2 on the basis of the gate shown in **b**. Less than 1.2% of the DO11.10 CD4 T cells in mice immunized 11 (**c**) or 60 (**d**) days earlier contained IL-2 or IFN- γ before challenge with ovalbumin peptide.

lymphoid tissues, might be improved by study of the lymphocytes that reside in non-lymphoid tissues. □

Methods

Adoptive transfer and flow cytometry

A sample of spleen and lymph node cells from OT-II TCR transgenic C57BL/6.PL mice¹⁰ was stained with phycoerythrin-labelled anti-Thy-1.1 antibody, CyChrome-labelled anti-CD4 antibody, and FITC-labelled anti-TCR V α 2 antibody (all from PharMingen), and analysed by flow cytometry. A volume of unlabelled cells containing $2.5\text{--}5 \times 10^6$ CD4⁺, TCR V α 2⁺ and Thy-1.1⁺ OT-II cells was then injected intravenously into each C57BL/6 recipient. We used this method to prepare the recipients used in the experiments shown in Figs 1a–c, 2 and 3. A similar method² was used in the experiments shown in Fig. 4 for transfer of spleen and lymph node cells from DO11.10 normal or Rag2^{−/−} TCR transgenic mice¹⁶, except that the transferred cells were identified with fluorochrome-labelled KJ1-26 anti-clonotypic antibody²⁷, and BALB/c mice were used as recipients. In the experiment shown in Fig. 1g, mixed bone marrow chimaeras containing roughly 1% DO11.10 T cells were produced by injecting 7×10^6 bone marrow cells from DO11.10 BALB/c severe combined immunodeficient (SCID) mice intravenously into irradiated (200 rads) BALB/c mice. We used chimaeras several months after irradiation. Naïve recipients were injected intravenously with 100 μ g ovalbumin peptide, 25–50 μ g LPS, or 100 μ g ovalbumin peptide plus 25–50 μ g LPS. In some cases, mice immunized 11 or 60 days previously with ovalbumin peptide plus LPS were challenged by i.v. injection of 100 μ g ovalbumin peptide. The cell-division history of DO11.10 T cells was determined by labelling the cells with 5,6-carboxy-fluorescein diacetate, succinimidyl ester (CFSE) before transfer, and measuring the loss of dye as described^{28,29}. We measured CD25 expression on DO11.10 T cells by flow cytometry²⁹.

Staining whole-mouse sections

Mice were killed and perfused before a popliteal lymph node was removed for flow cytometry. The mice or individual organs were then embedded in OCT medium, and frozen in liquid nitrogen. Ten-micron whole-body sections near the midline were prepared using an LKB 2250 cryomicrotome as described¹⁴. Sections were dehydrated in acetone for 10 min, and fixed in 1% formaldehyde for 15 min. Endogenous peroxidase activity was eliminated by incubating the sections in 1% H₂O₂ and 0.1% sodium azide for 1 h. Fc-binding sites and endogenous sources of biotin were blocked with 1% mouse serum, 1% rat serum, and anti-FcR monoclonal antibody (24G2, ATCC), followed by avidin, and then biotin solutions (Vector). Biotin-labelled anti-Thy-1.1 or KJ1-26 monoclonal antibody was added for 20 min. The sections were then incubated sequentially with streptavidin-peroxidase and biotinyl-tyramide from the TSA-Biotin kit (NEN) according to the manufacturer's instructions. The deposited biotin was detected with streptavidin-Cy3 (Caltag) and the sections were stained with a 10 μ g ml^{−1} solution of DAPI (Boehringer-Mannheim) in PBS. Each section was placed on a 7.3 × 10.2-cm glass plate, bathed in mounting medium (Vector), and covered with a coverslip.

Two sets of ×40 images covering the entire surface of each section were acquired with a CCD camera attached to an Olympus fluorescence microscope equipped with an automated stage and driven by Metamorph software (Universal Imaging Corporation), one set in the DAPI channel and another in the Cy3 channel. Photoshop 5.5 software (Adobe) was used to assemble the roughly 700 individual DAPI and Cy3 images into a single image for each section. This was accomplished by first producing 20 rows, each consisting of 35 adjacent images, which were then stacked sequentially to form the final image. The red objects on the assembled Cy3 image were enlarged fivefold so that individual T cells could be seen at the resulting low magnification. The DAPI and Cy3 images were overlaid to give the final dual-colour composite images shown in Fig. 1.

Quantification of transferred T cells

We quantified the number of OT-II cells within a given organ using Photoshop 5.5 software from representative ×40 images. The background signal in each organ was blanked (see below). A Cy3 image from an organ from a negative control animal that did not receive OT-II cells was converted to greyscale. We selected an area that lacked cell-sized objects. The threshold level within that area was adjusted so that no pixels were detected. We then applied this threshold level to the entire image. The number of pixels remaining in the image was then determined using the colour range function, and divided by the total number of pixels in that image. The same set of operations was performed on an image of the same organ from an animal that received OT-II cells. The fractional area from the negative control animal was subtracted from the area obtained from the OT-II recipient. The resulting value was multiplied by the volume of that organ (as determined by the weight of that organ in comparably treated animals) and divided by the average volume of a T cell (determined from the average diameter of stained OT-II cells from the spleen of the corresponding section), yielding the total number of OT-II cells per organ. The limit of detection was calculated by multiplying the average fractional area of an image occupied by a single T cell, by the volume of that organ, and divided by the average volume of a T cell. Lymph nodes and Peyer's patches could not be consistently analysed on whole-body sections because the plane of many individual sections did not, by chance, contain these tissues. For this reason, the number of OT-II cells in lymph nodes (and Peyer's patches) was extrapolated from the percentage of CD4⁺, TCR V α 2⁺ and Thy 1.1⁺ OT-II cells obtained by flow cytometric analysis of the explanted popliteal lymph nodes, on the basis of the knowledge that the T-cell response after i.v. injection occurs equally in all lymphoid tissues⁵.

Intracellular staining

BALB/c recipients of splenic and lymph node DO11.10 CD4 T cells were injected intravenously with ovalbumin peptide plus LPS. Eleven or sixty days later, some mice were challenged intravenously with ovalbumin peptide. Mice were killed 2 h later and perfused before lungs, spleen and lymph nodes were removed. All tissues were minced and incubated in collagenase D (Sigma, 400 U ml^{−1}) for 20 min at 37 °C in medium supplemented with 10 μ g ml^{−1} brefeldin A (Calbiochem) to prevent lymphokine secretion. Cells were fixed in 2% formaldehyde, permeabilized with 0.3% saponin, and stained with fluorochrome-labelled anti-CD4, KJ1-26 and anti-IL-2 or anti-IFN- γ antibodies to detect lymphokine production by DO11.10 T cells, as described²³.

Received 18 December 2000; accepted 12 January 2001.

- Kyburz, D. *et al.* T cell immunity after a viral infection versus T cell tolerance induced by soluble viral peptides. *Eur. J. Immunol.* **23**, 1956–1962 (1993).
- Kearney, E. R., Pape, K. A., Loh, D. Y. & Jenkins, M. K. Visualization of peptide-specific T cell immunity and peripheral tolerance induction *in vivo*. *Immunity* **1**, 327–339 (1994).
- McHeyzer-Williams, M. G. & Davis, M. M. Antigen-specific development of primary and memory T cells *in vivo*. *Science* **268**, 106–111 (1995).
- Murali-Krishna, K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
- Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relation. *Science* **272**, 54–60 (1996).
- Dutton, R. W., Bradley, L. M. & Swain, S. L. T cell memory. *Annu. Rev. Immunol.* **16**, 201–223 (1998).
- Mackay, C. R. Homing of naive, memory and effector lymphocytes. *Curr. Opin. Immunol.* **5**, 423–427 (1993).
- Butcher, E. C. & Picker, L. J. Lymphocyte homing and homeostasis. *Science* **272**, 60–66 (1996).
- Pape, K. A. *et al.* Use of adoptive transfer of T-cell-antigen-receptor-transgenic T cells for the study of T-cell activation *in vivo*. *Immunol. Rev.* **156**, 67–78 (1997).
- Barnden, M. J., Allison, J., Heath, W. R. & Carbone, F. R. Defective TCR expression in transgenic mice constructed using cDNA-based alpha- and beta-chain genes under the control of heterologous regulatory elements. *Immunol. Cell. Biol.* **76**, 34–40 (1998).
- Dresser, D. W. Specific inhibition of antibody production. II. Paralysis induced in adult mice by small quantities of protein antigen. *Nature* **191**, 1169–1171 (1962).
- Chiller, J. M., Habicht, G. S. & Weigle, W. O. Kinetic differences in unresponsiveness of thymus and bone marrow cells. *Science* **171**, 813–815 (1971).
- Chiller, J. M. & Weigle, W. O. Termination of tolerance to human gamma globulin in mice by antigen and bacterial lipopolysaccharide (endotoxin). *J. Exp. Med.* **137**, 740–750 (1973).
- Southern, P. J., Blount, P. & Oldstone, M. B. Analysis of persistent virus infections by *in situ* hybridization to whole-mouse sections. *Nature* **312**, 555–558 (1984).
- Pape, K. A., Khoruts, A., Mondino, A. & Jenkins, M. K. Inflammatory cytokines enhance the *in vivo* clonal expansion and differentiation of antigen-activated CD4⁺ T cells. *J. Immunol.* **159**, 591–598 (1997).
- Murphy, K. M., Heimberger, A. B. & Loh, D. Y. Induction by antigen of intrathymic apoptosis of CD4⁺CD8⁺ thymocytes *in vivo*. *Science* **250**, 1720–1723 (1990).
- Pape, K. A., Merica, R., Mondino, A., Khoruts, A. & Jenkins, M. K. Direct evidence that functionally impaired CD4⁺ T cells persist *in vivo* following induction of peripheral tolerance. *J. Immunol.* **160**, 4719–4729 (1998).
- Bradley, L. M. & Watson, S. R. Lymphocyte migration into tissue: the paradigm derived from CD4 subsets. *Curr. Opin. Immunol.* **8**, 312–320 (1996).
- Alferink, J. *et al.* Control of neonatal tolerance to tissue antigens by peripheral T cell trafficking. *Science* **282**, 1338–1341 (1998).
- Bell, E. B. & Sparshott, S. M. Interconversion of CD45R subsets of CD4 T cells *in vivo*. *Nature* **348**, 163–166 (1990).
- Westermann, J. & Pabst, R. How organ-specific is the migration of 'naïve' and 'memory' T cells? *Immunol. Today* **17**, 278–282 (1996).
- Vella, A. T., McCormack, J. E., Linsley, P. S., Kappler, J. W. & Marrack, P. Lipopolysaccharide interferes with the induction of peripheral T cell death. *Immunity* **2**, 261–270 (1995).
- Khoruts, A., Mondino, A., Pape, K. A., Reiner, S. L. & Jenkins, M. K. A natural immunological adjuvant enhances T cell clonal expansion through a CD28-dependent, interleukin (IL)-2-independent mechanism. *J. Exp. Med.* **187**, 225–236 (1998).
- Kim, S. K. *et al.* Generation of mucosal cytotoxic T cells against soluble protein by tissue-specific environmental and costimulatory signals. *Proc. Natl Acad. Sci. USA* **95**, 10814–10819 (1998).
- Kim, S. K., Schluns, K. S. & Lefrançois, L. Induction and visualization of mucosal memory CD8 T cells following systemic virus infection. *J. Immunol.* **163**, 4125–4132 (1999).
- Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).
- Haskins, K. *et al.* The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. *J. Exp. Med.* **157**, 1149–1169 (1983).
- Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**, 131–137 (1994).
- Merica, R., Khoruts, A., Pape, K. A., Reinhardt, R. L. & Jenkins, M. K. Antigen-experienced CD4 T cells display a reduced capacity for clonal expansion *in vivo* that is imposed by factors present in the immune host. *J. Immunol.* **164**, 4551–4557 (2000).

Acknowledgements

We thank J. Walter for technical assistance; D. Mueller, S. Jameson, M. Mescher, A. Haase, P. Southern, R. A. Reinhardt and K. Hogquist for discussions; L. Lefrançois for supplying OT-II mice; and Gerald Sedgewick for help with image analysis. Supported by grants from the NIH (M.K.J., A.K., R.M. and R.L.R.), the Cancer Research Institute (T.Z.), and the Howard Hughes Medical Institute (A.K.).

Correspondence and requests for materials should be addressed to M.K.J. (e-mail: marcj@mail.ahc.umn.edu).

Skewed maturation of memory HIV-specific CD8 T lymphocytes

Patrick Champagne^{*†‡}, Graham S. Ogg^{‡§}, Abigail S. King[§], Christian Knabenhans^{*}, Kim Ellefsen^{*}, Massimo Nobile^{*}, Victor Appay[§], G. Paolo Rizzardi[†], Sylvain Fleury^{*}, Martin Lipp^{||}, Reinhold Förster^{||}, Sarah Rowland-Jones[§], Rafick-P. Sékaly^{†¶}, Andrew J. McMichael[§] & Giuseppe Pantaleo^{*}

^{*} Laboratory of AIDS Immunopathogenesis, Divisions of Immunology and Allergy and Infectious Diseases, Department of Medicine, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

[†] Laboratoire d'Immunologie, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, H2W 1T8 Montréal and Department of Medicine, Division of Experimental Medicine, McGill University, H3A 2T5 Montréal, Canada

[§] MRC Human Immunology Unit, Institute for Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK

^{||} Molecular Tumor genetics and Immunogenetics, Max-Planck-Center for Molecular Medicine, 13092 Berlin, Germany

[¶] Department of Microbiology and Immunology, McGill University, H3A 2T5 Montréal and Département de Microbiologie et Immunologie, Université de Montréal, H2W 1T8 Montréal, Canada

[‡] These authors contributed equally to this work

Understanding the lineage differentiation of memory T cells is a central question in immunology. We investigated this issue by analysing the expression of the chemokine receptor CCR7, which defines distinct subsets of naive and memory T lymphocytes with different homing and effector capacities^{1–3} and antiviral immune responses to HIV and cytomegalovirus. *Ex vivo* analysis of the expression of CD45RA and CCR7 antigens, together with *in vitro* analysis of the cell-division capacity of different memory CD8⁺ T-cell populations, identified four subsets of HIV- and CMV-specific CD8⁺ T lymphocytes, and indicated the following lineage differentiation pattern: CD45RA⁺CCR7⁺ → CD45RA⁺CCR7[−] → CD45RA[−]CCR7[−] → CD45RA[−]CCR7⁺. Here we demonstrate through analysis of cell division (predominantly restricted to the CCR7⁺CD8⁺ T-cell subsets) that the differentiation of antigen-specific CD8⁺ T cells is a two-step process characterized

initially by a phase of proliferation largely restricted to the CCR7⁺CD8⁺ cell subsets, followed by a phase of functional maturation encompassing the CCR7[−]CD8⁺ cell subsets. The distribution of these populations in HIV- and CMV-specific CD8⁺ T cells showed that the HIV-specific cell pool was predominantly (70%) composed of pre-terminally differentiated CD45RA[−]CCR7[−] cells, whereas the CMV-specific cell pool consisted mainly (50%) of the terminally differentiated CD45RA⁺CCR7[−] cells. These results demonstrate a skewed maturation of HIV-specific memory CD8⁺ T cells during HIV infection.

A vigorous HIV-specific CD8⁺ T-cell immune response is readily detected during primary infection⁴, and the role of CD8⁺ T cells in the control of virus replication⁵ is well demonstrated; however, this control is incomplete and progressively lost as disease progresses. This has led to the hypothesis that the HIV-specific cytotoxic response is only partially effective⁶.

We analysed blood mononuclear cells obtained from 18 HIV-infected subjects with no previous anti-retroviral therapy with monoclonal antibodies to CD8 and CCR7, and with HLA tetrameric complexes loaded with the relevant peptides that allow direct evaluation of antigen-specific memory CD8⁺ T cells^{7,8}. Most (70–80%) of the CD8⁺ T cells were CCR7[−] in patients 1,020 and 1,021 (Fig. 1a), and these results were confirmed in a further 16 patients (mean 76 ± 13.5%). The immunodominant epitope SLYNTVATL⁹ in the gag protein, which is restricted by HLA-A2, was recognized by patients 1,020 and 2,121 (0.77 and 0.46% of blood mononuclear cells, respectively) (Fig. 1a). Most of the HIV-specific tetramer⁺(Tet⁺) cells (~70–80%) were contained in the CCR7[−] population (Fig. 1a). The prevalence of HIV-specific Tet⁺ CD8⁺ cells within the CCR7[−] cell population was confirmed in a larger number (*n* = 16) of patients. The CCR7[−]Tet⁺CD8⁺/CCR7⁺Tet⁺CD8⁺ cell ratio was 2.4/1.

Although most of the antigen-primed memory CD8⁺ T lymphocytes are phenotypically characterized by the expression of the CD45RO isoform¹⁰, a fraction of memory CD8⁺ T cells express the CD45RA isoform^{11,12}. We then determined whether the distribution of CCR7 could further identify, *in vivo*, other subsets of CD8⁺ T cells. In 3 (out of 5) representative patients most (70–80%) of the HIV-specific CD8⁺Tet⁺ cells were contained within the CD45RA[−] cell subset (Fig. 1b). More than 90% of the Tet⁺-CD45RA[−] cells were CCR7[−], whereas only a small percentage (5–10%) were CCR7⁺ (Fig. 1b). About 70–80% of Tet⁺CD45RA⁺ cells

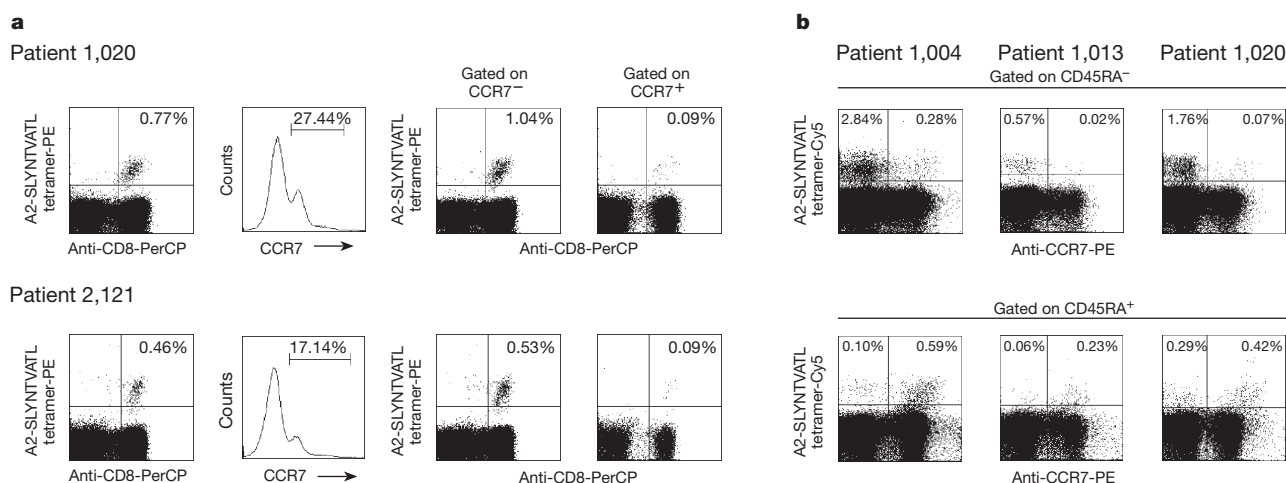


Figure 1 Distribution of CCR7 in different subsets of HIV-specific CD8⁺Tet⁺ cells. **a**, Blood mononuclear cells stained with anti-CD8 and anti-CCR7 monoclonal antibodies and with the A2-SLYNTVATL tetramer. Most CD8⁺ T cells as well as most HIV-specific Tet⁺ cells are contained within the CCR7[−] cell subset. **b**, Blood mononuclear cells stained with anti-CD8, anti-CD45 and anti-CCR7 monoclonal antibodies and with the A2-SLYNTVATL

tetramer. Upon analysis on gated CD45RA[−] and CD45RA⁺ T-cell populations, four subsets of HIV-specific CD8⁺ cells were identified. Most CD8⁺Tet⁺ cells are CCR7[−] within the CD45RA[−] cell subset, whereas they are CCR7⁺ within the CD45RA⁺ cell subset. CD8⁺Tet⁺ cells are predominantly contained within the CD45RA[−] cell subset. For all the flow cytometric analyses (Figs 1b, 2b, 3a–c and 4a–c) 1.5 × 10⁶ events were accumulated.

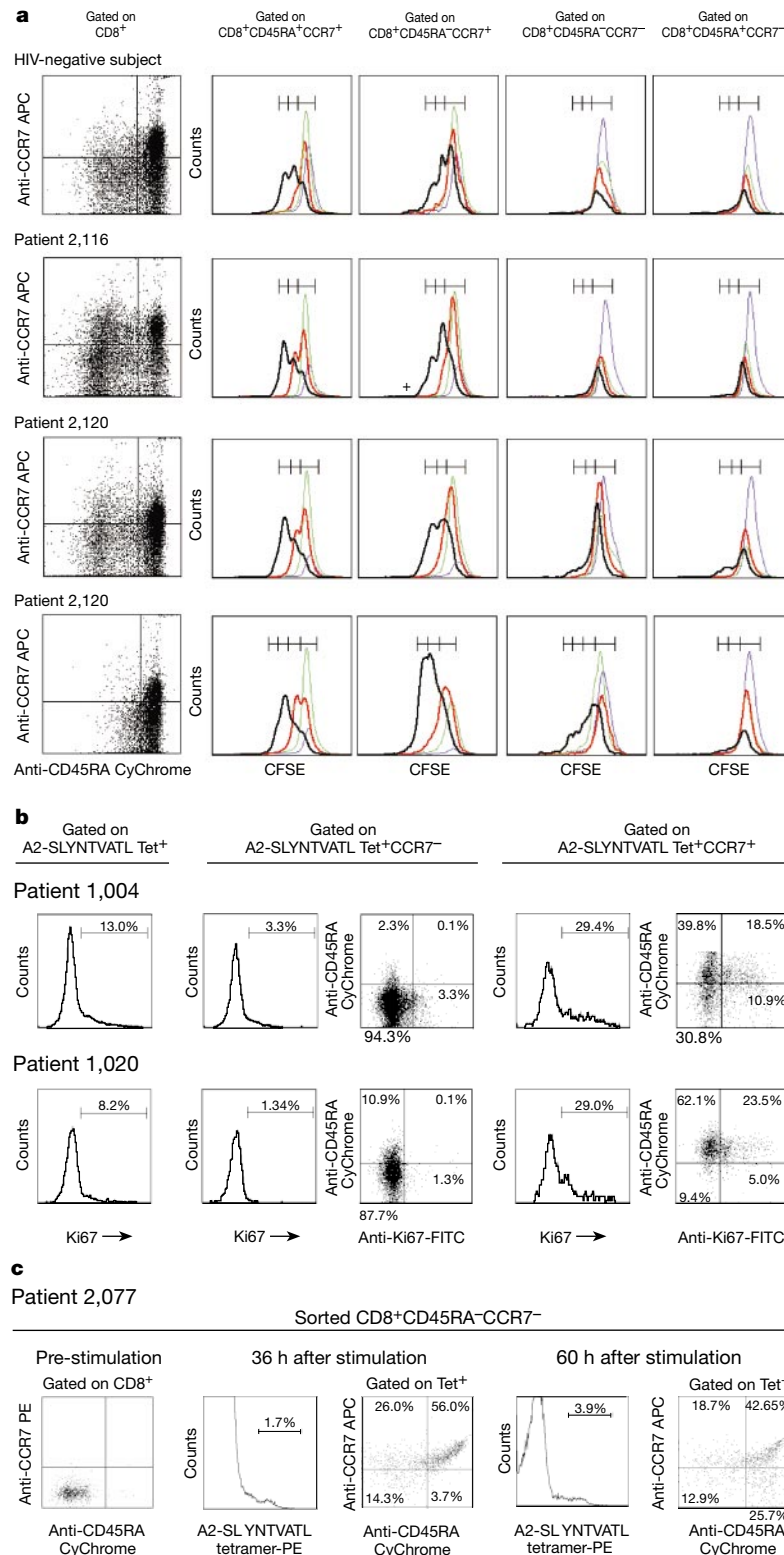


Figure 2 Proliferative capacity and differentiation pattern of different subsets of CD8⁺ T lymphocytes. **a**, Unfractionated blood mononuclear cells were labelled with CFSE, stimulated with anti-CD3 plus anti-CD28 monoclonal antibody. The distribution of CCR7 and CD45RA on gated CD8⁺ T cells is shown. Cell division was determined in different subsets of CD8⁺ T lymphocytes identified by the expression of CD45RA and CCR7 at different time points. Peaks correspond to each division cycle. Peaks at baseline (blue), 36 h (green), 48 h (red) and 60 h (black) are shown. **b**, *Ex vivo* analysis of cell division in Tet⁺CCR7⁺ and Tet⁺CCR7⁺ cell populations. Blood mononuclear cells were first stained for expression of surface markers and then for intracellular expression of Ki67 nuclear

antigen. On analysis of gated Tet⁺CCR7⁺ and Tet⁺CCR7⁺ cell populations, expression of Ki67 was mostly restricted to the Tet⁺CCR7⁺ cell subset. **c**, Analysis of the differentiation pattern of different CD8⁺ cell subsets after peptide HIV-specific stimulation *in vitro*. Blood mononuclear cells were stained with anti-CD8, anti-CD45RA and anti-CCR7 monoclonal antibodies, and sorted for CD8⁺CD45RA⁺CCR7⁺ cells. The sorted cell population was then stimulated with the relevant peptide, cultured in the presence of IL-2, and analysed at 36 and 60 h for the expression of CD45RA, CCR7, and for the presence of Tet⁺ cells. The purity of the sorted populations was greater than 96%. The experiments shown in **c** are representative of two separate experiments.

expressed CCR7, whereas 10–20% of cells were CCR7⁺ (Fig. 1b). Four cell populations of HIV-specific CD8⁺ T cells have been identified: 1) CD45RA⁺CCR7⁺; 2) CD45RA⁺CCR7⁺; 3) CD45RA⁺CCR7⁺; and 4) CD45RA⁺CCR7⁺. On the basis of their high frequency, it is clear that the CD45RA⁺Tet⁺CD8⁺ T cells are antigen-primed and expanded cells. Furthermore, the CD45RA⁺CCR7⁺ cells represented 15 ± 5.2% (mean ± s.d.) of the total HIV-specific Tet⁺ cells, the CD45RA⁺CCR7⁺ cells represented 4.1 ± 3.4%, the CD45RA⁺CCR7⁺ cells represented 71.8 ± 5.8% and the CD45RA⁺CCR7⁺ cells represented 8.9 ± 9.2%.

To determine the lineage differentiation pattern of these four subsets of CD8⁺ T lymphocytes, we first analysed the capacity of cell division for the four cell subsets on *in vitro* stimulation, and then the differentiation patterns of purified CD8⁺CCR7⁺CD45RA⁺ and CD8⁺CCR7⁺CD45RA⁺ cell populations after stimulation with the specific peptide. Unfractionated mononuclear cells isolated from

HIV-negative and HIV-infected subjects were stimulated with anti-CD3 plus anti-CD28 monoclonal antibody, and cell division was analysed by 5- (and 6-)carboxyl-fluorescein diacetate, succinimidyl ester (CFSE) labelling¹³ and flow cytometry in different subsets of CD8⁺ T cells characterized by the expression of CD45RA and CCR7. As shown for one (out of three) representative HIV-negative and three HIV-infected subjects, cell division was mostly confined to the populations of CD8⁺ T lymphocytes that express CCR7 (Fig. 2a). Furthermore, the appearance of cell division was more rapid in CD8⁺CD45RA⁺CCR7⁺ cells (48 h after stimulation) compared with CD8⁺CD45RA⁺CCR7⁺ cells (60 h after stimulation) (Fig. 2a). No evidence for significant cell division was observed in CD8⁺CD45RA⁺CCR7⁺ or in CD8⁺CD45RA⁺CCR7⁺ cells during the time of observation (Fig. 2a). No cell division was observed in any CD8⁺ T-cell subset in the unstimulated controls throughout the period of observation (data not shown).

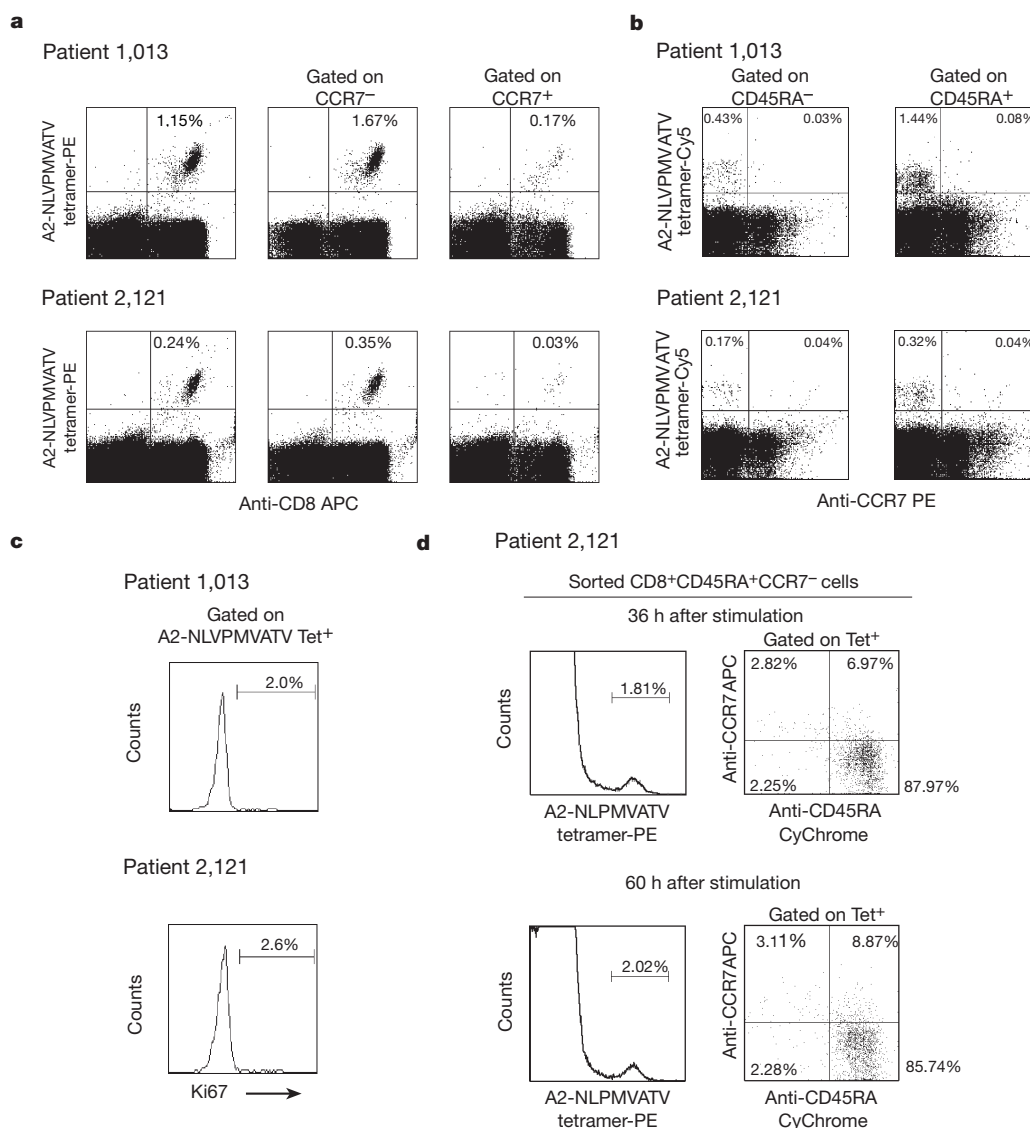


Figure 3 Distribution of CCR7 in different subsets of CMV-specific CD8⁺Tet⁺ cells.

a, Blood mononuclear cells were stained with anti-CD8 and anti-CCR7 monoclonal antibodies and with the A2-NLPMVATV tetramer. Most CMV-specific Tet⁺ cells are contained within the CCR7⁺ cell subset. **b**, Blood mononuclear cells were stained with anti-CD8, anti-CD45 and anti-CCR7 monoclonal antibodies and with the A2-NLPMVATV tetramer. On analysis within CD45RA⁺ and CD45RA⁺ T-cell subsets, four subsets of CMV-specific CD8⁺ cells were identified. Most CD8⁺Tet⁺ cells are CCR7⁺ within both the

CD45RA⁺ and CD45RA⁺ cell subsets, and are predominantly contained within the CD45RA⁺ cell subset. **c**, Expression of Ki67 on gated CMV-specific Tet⁺ cells. **d**, Analysis of phenotypic changes in the advanced differentiated CD8⁺CD45RA⁺CCR7⁺ cell population. The sorted CD8⁺CD45RA⁺CCR7⁺ cell population was stimulated by the CMV-specific peptide (see Methods), and the phenotypic changes were analysed at different time points. The purity of the sorted populations was greater than 98%. The experiments shown in **d** are representative of two separate experiments.

We then determined *ex vivo* the cell division capacity of HIV-specific CD8⁺ T-cell populations by intracellular staining for Ki67, a nuclear antigen that is expressed in all phases of the cell cycle except for the G0 phase¹⁴. About 10% of Tet⁺ cells were Ki67⁺ (Fig. 2b) in patients 1,004 and 1,020. Notably, only a small percentage (1–5%) of the Tet⁺CCR7⁺ cells expressed Ki67 compared with about 30% of Tet⁺CCR7⁺ cells (Fig. 2b). Tet⁺Ki67⁺ were present in both CD45RA⁺ and CD45RA⁺ cell subsets (Fig. 2b). Similar results were obtained in a further three patients. These results indicate that the cell division capacity is predominantly restricted to the CD8⁺ T-cell compartment that expresses CCR7. The low percentage of Ki67⁺ cells in the CCR7⁺ cell subset, which contains most of the Tet⁺ cells, is consistent with a reduced proliferative capacity of the HIV-specific cytotoxic T lymphocytes^{15,16}. Taken together, these results indicate that CCR7⁺CD8⁺ T cells—in particular CD8⁺CD45RA⁺CCR7⁺ cells—function as precursors of the CCR7⁺CD8⁺ T cells.

To assess further the lineage differentiation pattern of antigen-specific cells, we isolated CD45RA⁺CCR7⁺CD8⁺ cells by cell sorting, and cell cultures were phenotypically characterized at different time points after stimulation with the specific peptide and in the presence of interleukin-2 (IL-2). In patient 2,077, the percentage of Tet⁺ cells within the CD8⁺CD45RA⁺CCR7⁺ cell subset was 1.8% before stimulation (data not shown). After 36 h of peptide-specific stimulation there was no evidence for expansion of Tet⁺ cells (1.7%) within CD45RA⁺CCR7⁺CD8⁺ cells (Fig. 2c), and no substantial changes were observed at 60 h after stimulation (Fig. 2c). The lack of expansion of Tet⁺ cells within the stimulated CD45RA⁺CCR7⁺CD8⁺ cell population is consistent with the CFSE and Ki67 results, thus providing further evidence that this cell population cannot be efficiently expanded. These results further confirm the dichotomy with regard to the cell division and proliferative capacities among the CCR7⁺ and CCR7⁺CD8⁺ T-cell compartments: high cell division capacity/expansion of the CCR7⁺ versus limited capacity of cell division/expansion of the CCR7⁺CD8⁺ T-cell populations.

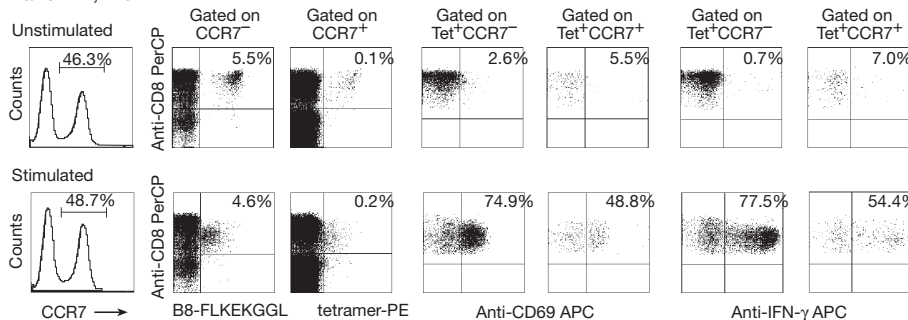
We next analysed the changes occurring in the expression of CD45RA and CCR7 upon *in vitro* antigen-specific stimulation of CD45RA⁺CCR7⁺CD8⁺ cells. A substantial proportion (26%) of Tet⁺

cells became CD45RA⁺CCR7⁺ (the upregulation of CCR7 after stimulation is consistent with previous studies¹⁷) whereas most (56%) acquired the CD45RA⁺CCR7⁺ phenotype at 36 h after stimulation (Fig. 2c). By 60 h after stimulation, a significant percentage (25%) of Tet⁺ cells exhibited the CD45RA⁺CCR7⁺ phenotype (Fig. 2c). No phenotypic changes were observed in the Tet⁺ cell populations from the stimulated cultures (data not shown).

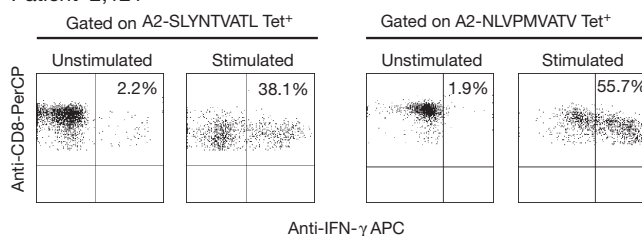
These results strongly suggest that CD45RA⁺CCR7⁺CD8⁺ T cells function as precursors for CD45RA⁺CCR7⁺CD8⁺, CD45RA⁺CCR7⁺CD8⁺ and CD45RA⁺CCR7⁺CD8⁺ cells. The results obtained from stimulated CD45RA⁺CCR7⁺ cells indicate that antigen-specific CD8⁺ T cells at an advanced stage of differentiation may also revert to a phenotype, such as CCR7⁺, which is typical of cells at earlier stages of cell differentiation, although they do not acquire the functional features, such as the proliferative capacity, of precursor cells.

Patients 1,013 and 2,121, in addition to the HLA-A2-restricted SLYNTVATL gag epitope (Fig. 1a and b), recognized the NLVPMVATV epitope¹⁸ in the pp65 protein of Cytomegalovirus (CMV) that is also restricted by HLA-A2. This provided an opportunity to determine the composition of the CMV-specific CD8⁺ T-cell pool. The percentage of CMV-specific Tet⁺CD8⁺ T cells was 1.15% and 0.24% in patients 1,013 and 2,121, respectively (Fig. 3a). About 70% of CMV-specific CD8⁺Tet⁺ cells were CCR7⁺ and 30% CCR7⁺ (Fig. 3a), whereas most of the CMV-specific Tet⁺ cells in both patients were CD45RA⁺CCR7⁺ (Fig. 3b). Similar analysis performed in a further six patients indicated that the CD45RA⁺CCR7⁺ cell subset represented about 5% (6 ± 1.5%; mean ± s.d.) of the total CMV-specific Tet⁺ cells, the CD45RA⁺CCR7⁺ represented 3.8 ± 1%, the CD45RA⁺CCR7⁺ represented 38.9 ± 6.4%, and the CD45RA⁺CCR7⁺ represented 51 ± 6.9% of the total Tet⁺ cells (Fig. 3b). Therefore, the CD45RA⁺CCR7⁺ (precursor cells) and the CD45RA⁺CCR7⁺CD8⁺ T-cell populations were expanded (up to 20% and 70% of Tet⁺ cells, respectively) in the HIV-specific but not in the CMV-specific cell pool. The CD45RA⁺CCR7⁺CD8⁺ T-cell subset was largely represented (50% of Tet⁺ cells) within the CMV-specific cell pool and poorly repre-

a Patient 2,123



b Patient 2,121



c

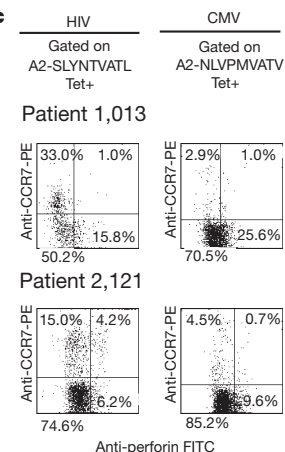


Figure 4 Functional analyses of HIV- and CMV-specific CD8⁺ T lymphocytes. **a**, Analysis of IFN-γ secretion in HIV-specific Tet⁺CCR7⁺ and Tet⁺CCR7⁺ cell subsets. On stimulation with the specific peptide, blood mononuclear cells were stained with the relevant tetramer and for the surface expression of CCR7, CD69, followed by intracellular staining for the detection of IFN-γ. A large percentage of cells expressed CD69 and stained positive for

IFN-γ within both Tet⁺CCR7⁺ and Tet⁺CCR7⁺ cell populations. **b**, Comparative analysis of IFN-γ secretion from HIV-specific and CMV-specific CD8⁺Tet⁺ cells after peptide-specific stimulation. **c**, Perforin expression in different subsets of HIV-specific and CMV-specific Tet⁺ cells. Most HIV-specific and CMV-specific Tet⁺ cells expressing perforin were contained within the CCR7⁺ cell subset.

sented within the HIV-specific CD8⁺ T-cell pool (5% of Tet⁺ cells). Furthermore, Ki67 nuclear antigen was expressed in only 2% of the CMV-specific Tet⁺ cells as compared with 10% of HIV-specific CD8⁺ T cells (Fig. 3c). As the proportion of Ki67⁺ cells can be used as a measure of T-cell production and turnover^{19,20}, these results indicate that there is at least a fivefold increase in the production of HIV-specific compared with CMV-specific CD8⁺ T cells, and suggest that there is a constant recruitment of T cells within the HIV-specific cell pool.

It has been proposed³ that the CD45RA⁺CCR7⁻ antigen-specific CD8⁺ cell population, which secretes interferon- γ (IFN- γ) and expresses high perforin levels, is composed of terminally differentiated CD8⁺ T cells. To address further this issue the CD45RA⁺CCR7⁻ CMV-specific CD8⁺ cell population was isolated by cell sorting, and cell cultures were analysed upon stimulation with the specific peptide. The percentage of Tet⁺ cells within the CD8⁺CD45RA⁺CCR7⁻ cell subset was 1.4% before stimulation (data not shown). No evidence for substantial expansion of CMV-specific Tet⁺ cells was observed at 36 (1.8%) and 60 h (2%) after peptide-specific stimulation, and there was no evidence for phenotypic changes (Fig. 3d). These results further support the hypothesis³ that the CD45RA⁺CCR7⁻CD8⁺ cells are at a terminal stage of differentiation. This, together with the different cell-division capacity of the different subsets of CD8⁺ T lymphocytes and the changes observed upon antigen-specific stimulation of CD45RA⁺CCR7⁻CD8⁺ cells, favours the following lineage differentiation pattern for antigen-specific T lymphocytes: CD45RA⁺CCR7⁺ → CD45RA⁻CCR7⁺ → CD45RA⁻CCR7⁻ → CD45RA⁺CCR7⁻.

The HIV- and CMV-specific CD8⁺ T cells were then functionally characterized for the production of IFN- γ and the expression of perforin. Patient 2,123 recognized the FLKEKGGGL epitope in the Nef protein restricted by HLA-B8 (Fig. 4a). Upon peptide-specific stimulation, a large percentage (50–75%) of CCR7⁺Tet⁺ and CCR7⁻Tet⁺ cells expressed CD69, a marker of activation, and stained positive for IFN- γ (Fig. 4a). Similar results were obtained in a further three patients (data not shown). These results indicate that not only CCR7⁻ memory CD8⁺ T cells, but also a large percentage of CCR7⁺ cells may secrete IFN- γ after antigen-specific stimulation.

We then compared the secretion of IFN- γ and the expression of perforin in two patients with HIV- and CMV-specific CD8⁺ T cells. We observed no differences in the secretion of IFN- γ (Fig. 4b). Perforin expression was reduced in HIV-specific CD8⁺ T cells of patient 1,013 and was predominantly confined to the CCR7⁻ cell subset (Fig. 4c). A defective expression of perforin²¹ within the HIV-specific Tet⁺CD8⁺ T cells compared with CMV-specific Tet⁺ cells⁶ has been shown. Our results suggest that this defective perforin expression could be related to the accumulation of T cells at a specific stage of differentiation rather than a defect in perforin production.

On the basis of the expression of CD45RA and CCR7, four subsets of memory CD8⁺ T lymphocytes have been identified, and their lineage differentiation pattern has been proposed (Fig. 5a). We have also observed large differences in the representation of the CD8⁺ T-cell subsets within the HIV- and CMV-specific memory CD8⁺ T-cell pools (Fig. 5b). The HIV-specific memory CD8⁺ T-cell pool is largely composed (70%) of the pre-terminally differentiated CD45RA⁻CCR7⁻ cells, whereas the terminally differentiated CD45RA⁺CCR7⁻ cell subset is poorly (less than 5%) represented. The latter, however, is the well represented (50%) cell subset within the CMV-specific memory CD8⁺ T-cell pool. The finding that most of the CMV-specific memory CD8⁺ T cells are terminally differentiated raises the issue of how it will be possible to generate a recall response to CMV if these cells are not able to proliferate upon antigen stimulation. In principle, the generation of a recall antigen response is thought to be associated with the proliferation of memory cells and their conversion to effector cells. Our results indicate a new model. In fact, the population of memory CMV-specific CD8⁺ T cells is not only composed of terminally differentiated effector (CD45RA⁺CCR7⁻) cells, but also of precursor (CD45RA⁺CCR7⁺) memory cells. It is therefore probable that the terminally differentiated effector cells are ready to rapidly intervene on antigen re-encounter, while precursor cells will expand and ensure continuous replenishment of the effector cell pool.

Taken together, our results demonstrate a skewed maturation of HIV-specific memory CD8⁺ T lymphocytes with the accumulation of a pre-terminally differentiated subset of memory cells. The

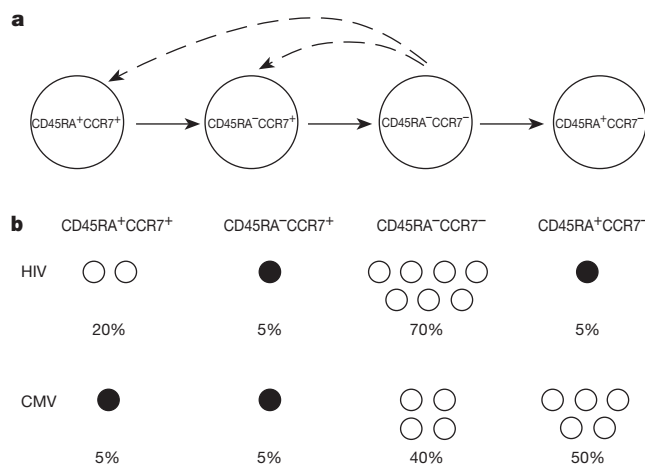


Figure 5 Lineage differentiation pattern of memory CD8⁺ T lymphocytes. **a**, The model is based on the identification of four subsets of memory CD8⁺ T lymphocytes characterized by the surface expression of CD8, CD45RA, CCR7 and by their proliferative capacity, and by the differentiation patterns of the different populations of memory CD8⁺ T lymphocytes observed after stimulation *in vitro*. CD8⁺CD45RA⁺CCR7⁺ cells function as precursors for the other subsets of memory cells. The CD8⁺CD45RA⁺CCR7⁻ cell population is the subset of memory CD8⁺ T cells at the most advanced stage of differentiation. Dashed arrows indicate that CD8⁺CD45RA⁻CCR7⁻ cells may revert to the phenotype of immediate precursor cells, that is, CD45RA⁻CCR7⁺, or even to CD45RA⁺CCR7⁺ cells upon antigen

stimulation. These phenotypic changes are not associated with the acquisition of functions typical of cells at early stages of differentiation. Functional characterization of the above subsets in HIV-specific and CMV-specific CD8⁺ T lymphocytes has shown that expression of perforin and IFN- γ secretion are associated with the late stages of differentiation, that is, predominantly to the CCR7⁻ cell subsets. **b**, Differences in the composition of the HIV- and CMV-specific memory CD8⁺ T-cell subsets. Each white cell corresponds to 10% of the corresponding CD8⁺ T-cell subset, whereas grey cells define cell subsets below 10%.

mechanisms responsible for these findings may include rapid consumption of terminally differentiated CD8⁺ T cells as a result of increased turnover²², high dose antigen-induced tolerance^{23,24}, and lack of antigen-specific CD4 helper activity^{25,26}. Based on our results, the high levels of antigen stimulation together with the lack of adequate HIV-specific CD4 helper activity^{25,26} may have kept HIV-specific memory CD8⁺ T lymphocytes at the stage of CD45RA⁺CCR7⁺ cells, and prevented their further differentiation to CD45RA⁺CCR7⁺ cells. These results provide new insights to the delineation of the memory CD8⁺ antiviral immune response, and for the evaluation of the immune response induced by new vaccine candidates. □

Methods

Patients

All patients in this study have been enrolled in two clinical therapeutic trials (CNAB2006 study and AVIB study). These trials are open-label, observational, non-randomized prospective studies carried out at a single site (Lausanne, Switzerland). Subjects included were HIV-1-infected therapy-naïve adults with a CD4 T-cell count ≥ 250 cells per μl and plasma viraemia $\geq 5,000$ HIV-1 RNA copies per ml. These studies were approved by the local Institutional Review Board, and all subjects gave written informed consent.

Tetrameric HLA molecules

We produced tetrameric HLA class I molecules as described^{7,8}. Prokaryotic expression vectors encoding for the extracellular portion of HLA-A2, -B7 and -B8, each tagged to a carboxy terminal BirA enzyme recognition sequence and $\beta 2$ microglobulin were separately transformed into *Escherichia coli*. Protein expression was induced with 1 mM isopropyl-thio- β -D-galactopyranoside and chains were extracted from inclusion bodies. Heavy and $\beta 2$ m chains were refolded by dilution in the presence of the appropriate peptides. For HLA-A2, HIV Gag p17 77–85 (SLYNTVATL) and Pol 476–484 (ILK-EPVHGV), CMV pp65 495–503 (NLVPMVATV); HLA-B7, HIV Nef 128–137 (TPGPGVRYPL), gp120 296–305 (RPNNTTRKSI) and Gag p24 (ATPQDLNTM); and HLA-B8, HIV Nef 89–97 (FLKEKGL), Gag p17 24–31 (GGKKKYKL) and Gag p24 259–267 (GEIYKRWII). Refolded complexes were enzymatically biotinylated with BirA (Avidity) and fast protein liquid chromatography purified by ion exchange on a MonoQ column or by size exclusion on a Sepharose 75 column (Pharmacia Biotech). Biotinylation efficiency was assessed by band shift of biotinylated monomer with streptavidin on an SDS-PAGE protein gel and routinely evaluated to 80–95%. Extravidin-phycoerythrin (PE) or Extravidin (Sigma) coupled to Cy5 (Amersham) were then mixed with biotinylated monomeric major histocompatibility complex to a 1/4 molar ratio.

FACS analysis and sorting

Cells cryo-preserved in liquid nitrogen were thawed for staining and analysis. Cells were either stained with PE-labelled tetramers for 15 min at 37 °C before 30 min incubation at 4 °C with other cell-surface antibodies, or with Cy5-coupled tetramers for 1 h at 4 °C with other surface-staining antibodies. Rat anti-human CCR7 antibody (3D12, rat IgG2a) staining was followed by goat anti-rat IgG(H + L)-fluorescein isothiocyanate (FITC), phycoerythrin (PE) (Southern Biotechnologies) or allophycocyanin (APC) (Caltag Laboratories). We used the following mouse anti-human antibodies in different combinations for cell-surface staining and sorting: anti-CD8 (IgG1, Leu-2a) coupled to FITC, PE, PerCp or APC and anti-CD45RO APC (IgG2a, UCHL-1) (Becton Dickinson), anti-CD45RA CyChrome (IgG2a, HI100) (PharMingen). For intracellular perforin and Ki67 analyses, after surface-marker labelling, cells were fixed and permeabilized with ORTHO PermeaFix (Ortho Diagnostic Systems) as per the manufacturer's instructions before intracellular staining with anti-perforin FITC (8G9, IgG2b) (PharMingen) or anti-Ki67 FITC (IgG1, MIB-1) (Immunotech), respectively. We used isotype-matched controls for intracellular stainings. Data was acquired on a FACSCalibur system, and analysed using CellQuest software (Becton Dickinson systems). We performed cell sorting on a FACS Vantage (Becton Dickinson systems).

In vitro stimulation and CFSE tracking of cell division

Peripheral blood mononuclear cells (PBMCs) were labelled with CFSE (Molecular Probes)¹³ by incubation with 0.6 μM CFSE for 10 min at 37 °C in PBS. Labelling was quenched with FCS and cells were washed in RPMI 10% FCS. CFSE-labelled cells were seeded at 10^6 cells per well in 48-well cell culture plates: for stimulated cultures, wells had been pre-coated overnight with 1 $\mu\text{g ml}^{-1}$ purified anti-CD3 (OKT3) in PBS and washed before plating of the cells. Wells for unstimulated control cultures had been pre-incubated with PBS alone. Co-stimulation was provided by soluble, purified anti-CD28 added at a final concentration of 1 $\mu\text{g ml}^{-1}$. FACS analysis of the cells was performed before culture and after 36, 48 and 60 h of *in vitro* culture.

Peptide-specific stimulation

PBMCs were thawed and incubated at 37 °C in RPMI 1640 Gutamax-1 medium (Gibco BRL, Life Technologies) containing 10% inactivated AB human serum (Sigma) (Rh 10). Autologous APCs were prepared from PBMCs that were washed, irradiated (3,000 rads) and plated at 2×10^5 cells per well in flat-bottom 96-well cell culture plates (Costar). These

APCs were incubated at 37 °C in RPMI 1640 Glutamax-1 medium for 3 h in the presence of 10 μM (final concentration) of the appropriate peptide. FACS sorted CD8⁺CD45RA⁺CCR7⁺ and CD8⁺CD45RA⁺CCR7⁺ cells were then seeded over the pulsed-irradiated autologous APCs at $1\text{--}2 \times 10^5$ cells per well in culture medium containing 10% human serum, and human IL-2 (Boehringer Mannheim, Germany) was added to 4 U ml^{-1} (final concentration). The T-cell populations were harvested and stained for FACS analysis at different time points after stimulation.

Cytokine detection

Intracellular cytokine production was assessed as described¹⁹. PBMCs were stimulated with 10 μM (final concentration) of specific peptide (or PBS for unstimulated controls) for 6 h at 37 °C, in the presence of 3 $\mu\text{g ml}^{-1}$ purified anti-CD28 antibody (Becton Dickinson) and, as of the second hour, with 10 $\mu\text{g ml}^{-1}$ Brefeldin A (Sigma). Cells to be activated were stained with tetramer-PE for 15 min at 37 °C before activation. Cell-surface staining was completed as described after the 6 h *in vitro* activation. Cells were then permeabilized with Orthopermeafix and labelled with anti-human IFN- γ APC (IgG1, B27) (PharMingen). Alternatively, activation was assessed by staining with anti-CD69 APC (Becton Dickinson).

Received 20 September 2000; accepted 16 January 2001.

- Butcher, E. C. & Picker, L. J. Lymphocyte homing and homeostasis. *Science* **272**, 60–66 (1996).
- Sallusto, F. *et al.* The role of chemokine receptors in primary, effector, and memory immune responses. *Annu. Rev. Immunol.* **18**, 593–620 (2000).
- Sallusto, F. *et al.* Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).
- Pantaleo, G. *et al.* Major expansion of CD8⁺ T cells with a predominant V β usage during the primary immune response to HIV. *Nature* **370**, 463–467 (1994).
- Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8⁺ lymphocytes. *Science* **283**, 857–860 (1999).
- Appay, V. *et al.* HIV-specific CD8⁺ T cells produce antiviral cytokines but are impaired in cytolytic function. *J. Exp. Med.* **192**, 63–75 (2000).
- Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
- Ogg, G. S. *et al.* Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103–2106 (1998).
- Johnson, R. P. *et al.* HIV-1 gag-specific cytotoxic T lymphocytes recognize multiple highly conserved epitopes. *J. Immunol.* **147**, 1512–1521 (1991).
- Michie, C. A. *et al.* Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264–265 (1992).
- Callan, M. F. C. *et al.* Direct visualization of antigen-specific CD8⁺ T cells during the primary immune response to Epstein-Barr Virus *in vivo*. *J. Exp. Med.* **187**, 1395–1402 (1998).
- Hamann, D. *et al.* Phenotypic and functional separation of memory and effector human CD8⁺ T cells. *J. Exp. Med.* **186**, 1407–1418 (1997).
- Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**, 131–137 (1994).
- Gerdes, J. *et al.* Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J. Immunol.* **133**, 1710–1715 (1984).
- Pantaleo, G. *et al.* Defective clonogenic potential of CD8⁺ T lymphocytes in patients with AIDS. Expansion *in vivo* of a nonclonogenic CD3⁺CD8⁺DR⁺CD25⁺ T cell population. *J. Immunol.* **144**, 1696–1704 (1990).
- Pantaleo, G. *et al.* CD8⁺ T lymphocytes of patients with AIDS maintain normal broad cytolytic function despite the loss of human immunodeficiency virus-specific cytotoxicity. *Proc. Natl Acad. Sci. USA* **87**, 4818–4822 (1990).
- Sallusto, F. *et al.* Switch in chemokine receptor expression upon TCR stimulation reveals novel homing potential for recently activated T cells. *Eur. J. Immunol.* **29**, 2037–2045 (1999).
- Wills, M. R. *et al.* The human cytotoxic T-lymphocyte (CTL) response to Cytomegalovirus is dominated by structural protein pp65: Frequency, specificity, and T-cell receptor usage of pp65-specific CTL. *J. Virol.* **70**, 7569–7579 (1996).
- Fleury, S. *et al.* Limited CD4⁺ T-cell renewal in early HIV-1 infection: effect of highly active antiretroviral therapy. *Nature Medicine* **4**, 794–801 (1998).
- Fleury, S. *et al.* Long-term kinetics of T cell production in HIV-infected subjects treated with highly active antiretroviral therapy. *Proc. Natl Acad. Sci. USA* **97**, 5393–5398 (2000).
- Andersson, J. *et al.* Perforin is not co-expressed with Granzyme A within cytotoxic granules in CD8⁺ T lymphocytes present in lymphoid tissue during chronic HIV infection. *AIDS* **13**, 1295–1303 (1998).
- Wolthers, K. C. *et al.* T cell telomere length in HIV-1 infection: no evidence for increased CD4⁺ T cell turnover. *Science* **274**, 1543–1547 (1996).
- Mitchison, N. A. *et al.* *Proc. R. Soc. Lond.* **161**, 275–293 (1964).
- Zinkernagel, R. M. Immunology taught by viruses. *Science* **271**, 173–178 (1996).
- Rosenberg, E. S. *et al.* Vigorous HIV-1-specific CD4⁺ T cell responses associated with control of viremia. *Science* **278**, 1447–1450 (1997).
- Miedema, F. *et al.* Changing virus-host interactions in the course of HIV-1 infection. *Immunol. Rev.* **140**, 35–72 (1994).

Acknowledgements

We thank A. Wilson for providing the Extravidin-Cy5 conjugate. This work was supported by an SNF grant (Tandem project), by the EuroVae project (G.P.) and by the Lenards Foundation (G.P.), and by an NIH grant (Acute Infection, G.P.; R.P.S.). P.C. is supported by a Doctoral Award of the Medical Research Council of Canada. R.P.S. is a Canadian Institutes of Health Research senior scientist.

Correspondence and requests for materials should be addressed to G.P. (e-mail: Giuseppe.Pantaleo@chuv.hospvd.ch).

A conserved XIAP-interaction motif in caspase-9 and Smac/DIABLO regulates caspase activity and apoptosis

Srinivasa M. Srinivasula*, Ramesh Hegde*, Ayman Saleh†, Pinaki Datta*, Eric Shiozaki‡, Jijie Chai‡, Ryung-Ah Lee*, Paul D. Robbins†, Teresa Fernandes-Alnemri*, Yigong Shi‡ & Emad S. Alnemri*

* Center for Apoptosis Research and the Department of Microbiology and Immunology, Kimmel Cancer Institute, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

† Department of Molecular Genetics and Biochemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA

‡ Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, New Jersey 08544, USA

X-linked inhibitor-of-apoptosis protein (XIAP) interacts with caspase-9 and inhibits its activity^{1–3}, whereas Smac (also known as DIABLO) relieves this inhibition through interaction with XIAP^{4–7}. Here we show that XIAP associates with the active caspase-9–Apaf-1 holoenzyme complex through binding to the amino terminus of the linker peptide on the small subunit of caspase-9, which becomes exposed after proteolytic processing of procaspase-9 at Asp 315. Supporting this observation, point mutations that abrogate the proteolytic processing but not the catalytic activity of caspase-9, or deletion of the linker peptide, prevented caspase-9 association with XIAP and its concomitant inhibition. We note that the N-terminal four residues of caspase-9 linker peptide share significant homology with the N-terminal tetra-peptide in mature Smac and in the *Drosophila* proteins Hid/Grim/Reaper^{8,9}, defining a conserved class of IAP-binding motifs. Consistent with this finding, binding of the caspase-9 linker peptide and Smac to the BIR3 domain of XIAP is mutually exclusive, suggesting that Smac potentiates caspase-9 activity by disrupting the interaction of the linker peptide of caspase-9 with BIR3. Our studies reveal a mechanism in which binding to the BIR3 domain by two conserved peptides, one from Smac and the other one from caspase-9, has opposing effects on caspase activity and apoptosis.

Recent data suggest that processing of procaspase-9 might be required for inhibition by XIAP³, as the overexpression of XIAP was not able to inhibit DNA damage-induced processing of procaspase-9 in U-937 cells, but inhibited the catalytic activity of processed caspase-9. To understand the mechanism of inhibition of the caspase-9–Apaf-1 holoenzyme complex, we decided to reconstitute *in vitro* caspase-9–Apaf-1 holoenzyme complexes containing either fully processed caspase-9 or unprocessed procaspase-9 and study their interaction with—and inhibition by—XIAP. To produce fully processed caspase-9 we overexpressed wild-type (WT) procaspase-9 in bacteria, which resulted in complete processing of procaspase-9 into its p35 and p12 subunits (Fig. 1b, lane 2). Sequence analysis of the purified recombinant caspase-9 revealed that >90% of caspase-9 processing in bacteria occurs at Asp 315, which generates the p35 and p12 subunits, and the remaining 10% of processing occurs at Asp 330 to generate the p10 subunit. Minor processing was also detected at Glu 306. To produce a recombinant unprocessed procaspase-9 it was necessary to mutate Asp 315, Asp 330 and Glu 306 to Ala (Fig. 1a). Overexpression of the triple-mutant procaspase-9 (E306/D315/D330A) produced an unprocessed protein (Fig. 1b, lane 3). When the processed WT caspase-9 or the triple-mutant procaspase-9 proteins were reconstituted with purified Apaf-1 at

physiological concentrations (20 nM each), the triple-mutant procaspase-9 was as efficient as the fully processed WT caspase-9 in processing procaspase-3 C163A (Fig. 1c). The triple mutant was also as efficient as the fully processed WT caspase-9 in inducing DEVD-AMC (Asp-Glu-Val-Asp-7-amino-4-methyl coumarin) cleavage in Apaf-1–caspase-9-deficient S100 fraction in the presence of Apaf-1, cytochrome c, and dATP, but not in their absence (Fig. 1d).

Given that both processed and unprocessed caspase-9–Apaf-1 holoenzyme complexes are catalytically active, we asked whether XIAP could inhibit them equally. As shown in Fig. 2a, XIAP did not significantly inhibit the processing of procaspase-3 by the holoenzyme that contains the mutant caspase-9 but completely inhibited the processing by the holoenzyme that contains the WT caspase-9. The loss of inhibition of the catalytic activity of the holoenzyme containing the mutant caspase-9 could be due to the inability of XIAP to associate with the uncleavable caspase-9 in the holoenzyme complex. To test this hypothesis, we analysed the two complexes after incubation with XIAP, using gel filtration on a Superose-6 fast protein liquid chromatography (FPLC) column. As shown in Fig. 2b, both WT and uncleavable caspase-9 can form large (molecular mass, 1.4×10^3 KDa) holoenzyme complexes with Apaf-1 after stimulation with cytochrome c and dATP. We note that XIAP co-migrated with the WT caspase-9–Apaf-1 complex but not with the uncleavable caspase-9–Apaf-1 complex (Fig. 2b). Activity assays on the peak fractions containing the holoenzymes revealed that the

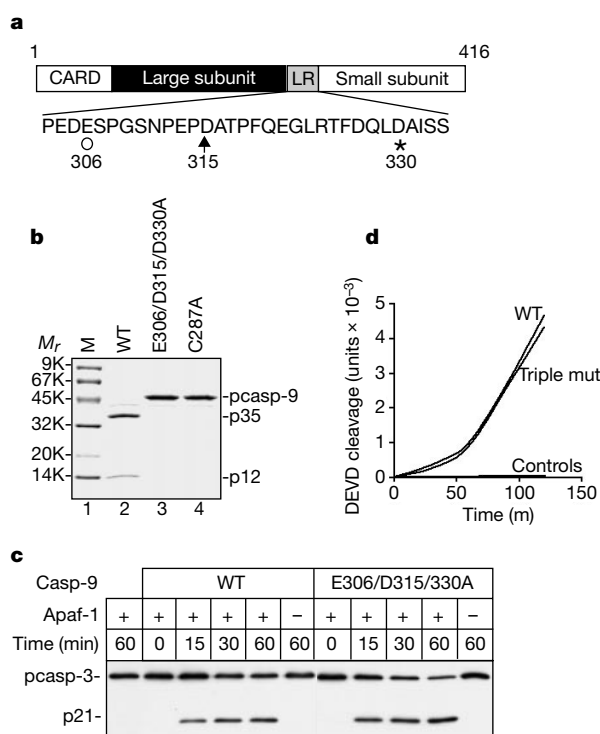


Figure 1 Proteolytic processing of procaspase-9 is not required for activation by Apaf-1. **a**, Bar diagram representing human procaspase-9 and the positions of the autocatalytic cleavage site (Asp 315, arrow) and caspase-3 cleavage site (Asp 330, star) within the linker region (LR) between the large and small subunits. The position of the minor autocatalytic cleavage site (Glu 306, circle) is also shown. **b**, Coomassie stained gel of recombinant WT (lane 2), triple-mutant E306/D315/D330A (lane 3) and active site mutant C287A caspase-9 (lane 4) purified on Talon-affinity resin from bacterial extracts. Lane 1, molecular-mass marker. **c**, Time course analysis of processing of procaspase-3 by the WT and triple-mutant caspase-9 holoenzyme complexes determined by western blotting. **d**, Time course analysis of the DEVD-AMC cleaving activity of caspase-9–Apaf-1-deficient S100 extracts reconstituted with Apaf-1 and WT or triple-mutant caspase-9 proteins. Controls represent WT and triple-mutant caspase-9 proteins incubated with the S100 extracts plus Apaf-1 without cytochrome c and dATP.

uncleavable complex was able to process procaspase-3 (Fig. 2c, lane II), whereas the WT complex was completely inactive (lane I). A control WT caspase-9–Apaf-1 complex that was reconstituted without XIAP was fully active (lane III). This shows that XIAP can associate with, and inhibit the activity of, the WT-caspase-9–Apaf-1 complex, but not the uncleavable caspase-9–Apaf-1 complex, suggesting that processing of caspase-9 at the interdomain linker region is important for binding to XIAP. To further confirm the gel filtration data, we incubated XIAP with purified WT or uncleavable caspase-9-holoenzyme complexes and then immunoprecipitated the complexes with an anti Apaf-1 antibody. Only the WT caspase-9–Apaf-1 complex contained XIAP (Fig. 2d, upper panel). These data are consistent with recent observations which revealed that XIAP does not inhibit activation of procaspase-9 but inhibits the activity of the processed caspase-9 in cells undergoing apoptosis³.

Because the uncleavable caspase-9–Apaf-1 complex is catalytically active, the inability of XIAP to associate with it and inhibit its

activity suggests that the association between caspase-9 and XIAP does not require the active site cysteine, but probably involves residues exposed after autoprocessing of procaspase-9 at Asp315. Examination of the free N-terminus of the human, mouse and *Xenopus* p12 subunit of caspase-9, generated after autoprocessing at Asp 315¹⁰, revealed that they all contain a 4-residue motif similar to the BIR3-interaction motif present at the N-terminus of mature Smac (Fig. 3a). This motif also has significant homology to the IAP-interaction motif at the N-termini of the *Drosophila* proteins Hid, Reaper and Grim^{8,9} (Fig. 3a). To determine whether this conserved motif interacts with XIAP, we performed *in vitro* interaction assays with ³⁵S-labelled full-length XIAP, or isolated BIR3 domain of XIAP and GST fusion proteins of p12, p10 or the linker peptide (Fig. 3b). Both p12 and the linker peptide were able to interact with full-length XIAP as well as the isolated BIR3 domain of XIAP. However, the p10-GST fusion protein was able to interact only weakly (~50 to 100 times less) with the full-length XIAP or the BIR3 domain. This weak interaction is due to the conservation of the first two residues of the BIR3-binding motif on the N-terminus of p10 (human, Ala331-Ile332; mouse, Ala331-Val332). Single point mutation of these two residues completely abolished the weak interaction between XIAP/BIR3 and p10 (data not shown). As the BIR3 domain of XIAP is the domain that specifically targets caspase-9¹, this suggests that the interaction between the linker peptide of caspase-9 and the BIR3 domain is primarily responsible for this inhibition. The above results were further confirmed using far-western blot analysis with ³⁵S-labelled XIAP. XIAP was able to bind to the WT p12 subunit of caspase-9, p12-GST and Smac-GST bands, but not to variants of caspase-9 with Asp 315 to Ala mutation or p10-GST bands on the nitrocellulose filter (Fig. 3c). The absence of interaction between p10 and XIAP by far western suggests that the observed weak interaction (Fig. 3b) is not physiologically significant. These results indicate that cleavage of caspase-9 at Asp 315 exposes the XIAP-binding motif in the linker region allowing binding to the BIR3 domain of XIAP and concomitant inhibition of caspase-9 activity.

To determine the importance of the linker peptide for inhibition of caspase-9, we deleted residues 316–330 from caspase-9 and expressed it in bacteria. This mutant (with a deleted linker: 'Δlinker') was able to undergo complete processing to generate the p35 and p10 subunits (Fig. 3d, left panel). As expected, the deletion mutant significantly lost the ability to interact with and to be inhibited by BIR3 (Fig. 3d, middle and right panels). This confirms that the p10 subunit of caspase-9 is not the primary target of XIAP-inhibition, and that the linker peptide is required for binding to BIR3 and for inhibition of caspase-9 activity.

To determine the importance of the first two residues of caspase-9-p12, we mutated them to SG or GG. The AT/SG or AT/GG mutants were completely processed at Asp 315 to generate the p35 and p12 subunits (Fig. 3d, left panel, lanes SG and GG). Like the linker-deletion mutant, both the SG and GG point mutant caspase-9 significantly lost the ability to interact with and to be inhibited by BIR3 (Fig. 3d, middle and right panels). This observation indicates that the first two residues in the p12 subunit of caspase-9 are important for binding to BIR3 and inhibition.

As caspase-3 is not inhibited by BIR3 (concentration giving 50% inhibition, IC₅₀ > 400 nM), we asked whether substitution of the first four residues of caspase-3-p12 with AVPF, which are the same residues present at the N-terminus of caspase-9-p12, could allow binding to and inhibition by BIR3. As shown in Fig. 3e, mutation of the first two residues of caspase-3 to AV allowed weak binding to XIAP and inhibition by BIR3 (IC₅₀ ≈ 140 nM). Mutation of the first four residues to AVPF enhanced binding to XIAP and increased inhibition by BIR3 (IC₅₀ ≈ 4 nM). These mutations also enhanced inhibition of caspase-3 by BIR2 of XIAP (IC₅₀ values: WT, ~10 nM; AV, ~7 nM; AVPF, ~4 nM). This is consistent with the recent findings that BIR2 can also bind the AVPI peptide of Smac^{7,11}.

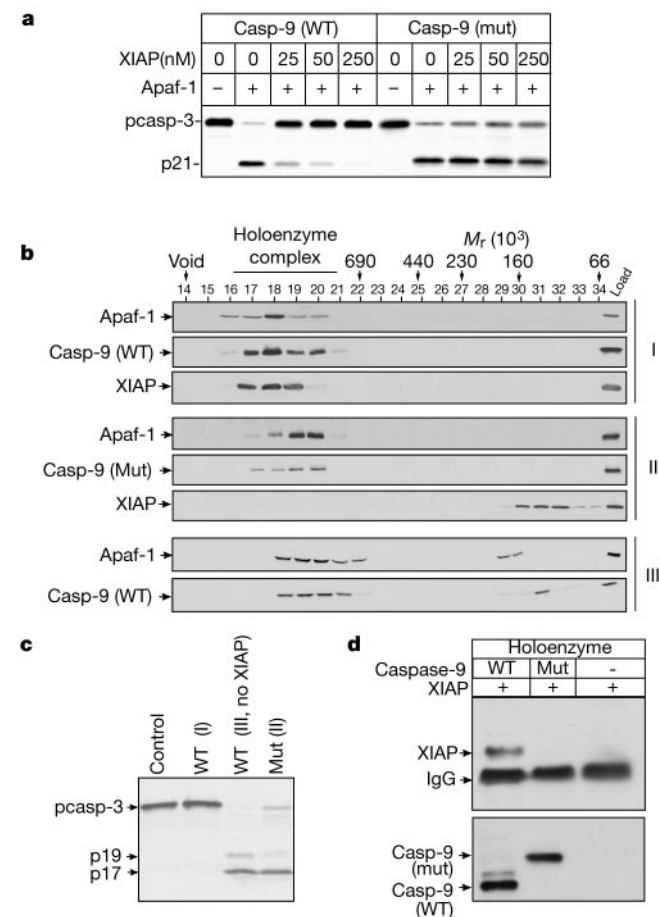


Figure 2 Proteolytic processing of caspase-9 is important for XIAP-inhibition. **a**, Autoradiographic analysis of processing of ³⁵S-labelled procaspase-3 by the fully processed WT, and the unprocessed triple-mutant caspase-9 holoenzyme complexes in the presence of different amounts of XIAP. **b**, Superose-6 gel filtration analysis of the WT or triple-mutant caspase-9 holoenzyme complexes reconstituted with XIAP (panels I and II) or nothing (panel III). The column fractions were analysed by western blotting with anti-Apaf-1, anti caspase-9 or anti-XIAP antibodies. **c**, Autoradiographic analysis of processing of ³⁵S-labelled procaspase-3 by the holoenzyme complexes from runs I (WT), III (WT, no XIAP) and II (Mut), respectively. Lane 1, buffer control. **d**, FPLC-purified holoenzyme complexes containing WT or triple-mutant caspase-9 were incubated with XIAP and then immunoprecipitated with anti-Apaf-1 antibody. The immunoprecipitates were analysed by western blotting with an XIAP antibody (upper panel) or a caspase-9 antibody (lower panel).

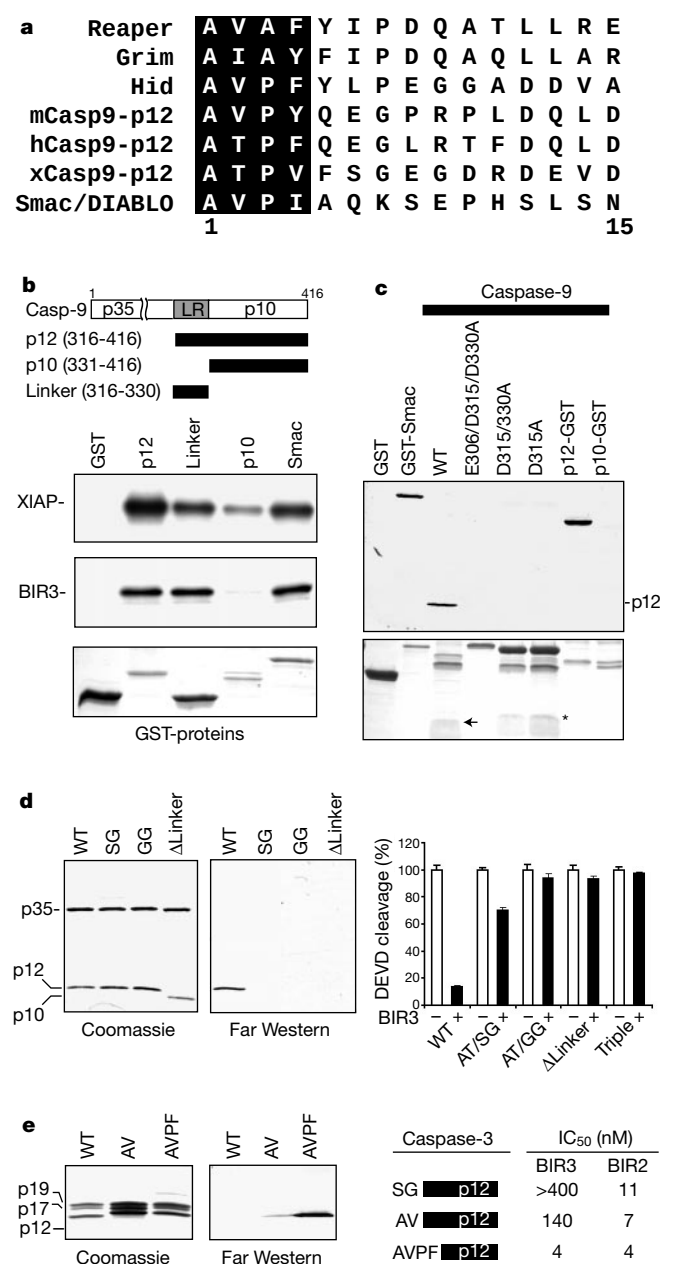


Figure 3 The N terminus of the p12 subunit of caspase-9 harbours a BIR3-binding motif. **a**, Collinear alignment of the N-terminal sequences of *Drosophila* Reaper, Grim and Hid, mouse, human and *Xenopus* caspase-9-p12, and human Smac. The BIR3-binding motif is highlighted. **b**, Autoradiographic analysis of the interactions of ³⁵S-labelled XIAP, or BIR3 domain of XIAP with C-terminal GST-tagged caspase-9-p12 (residues 316–416, p12), linker region (residues 316–330, linker) or p10 (residues 331–416, p10), or mature Smac (Smac). **c**, Upper panel, far-western blot analysis of recombinant WT or mutant caspase-9, or Smac-GST fusion protein with ³⁵S-labelled XIAP as described in Methods. Lower panel, Coomassie stained gel of all the indicated proteins. Arrow indicates p12 of caspase-9. Asterisk indicates p14 of caspase-9, which is generated by processing at E306. **d**, Equal amounts of recombinant WT or mutant caspase-9 (AT316, 317SG, AT316, 317GG and Δlinker) were fractionated by SDS-PAGE and Coomassie stained (left panel), or analysed by far western as in **c** (middle panel), or reconstituted with Apaf-1 and cytochrome c plus dATP in the presence (+) or absence (–) of purified XIAP-BIR3 (20 nM) in caspase-9–Apaf-1 deficient S100 extracts (right panel). The activity of each reconstituted extract is represented as % DEVD cleavage relative to the DEVD-AMC cleaving activity in the absence of BIR3. **e**, Recombinant WT or mutant caspase-3 (SG176, 177AV or SGVD176–179AVPF) were fractionated by SDS-PAGE and Coomassie stained (left panel) or analysed by far western as in **c** (middle panel). Right panel shows the calculated IC₅₀ values for the WT (SG) and the indicated caspase-3 mutants as calculated from the percentage of inhibition by BIR3-RING (BIR3) or BIR1-BIR2 (BIR2) of XIAP.

Together, the above results clearly establish that inhibition of human/mouse caspase-9 by XIAP is due to interaction of the ATPF/AVPY motif at the N-terminus of p12 with the BIR3 domain of XIAP.

Smac may promote caspase-9 activity by interfering with the interaction of the caspase-9-p12 with the BIR3 domain of XIAP. To determine if binding of caspase-9-p12 and Smac to the BIR3 domain is mutually exclusive, we performed *in vitro* binding experiments between Smac or caspase-9-p12 and BIR3 in the presence or absence of a chemically synthesized caspase-9 linker peptide or Smac-N7 peptide, respectively. As shown in Fig. 4a, the linker peptide completely inhibited the binding of Smac to BIR3. Similarly, the Smac-N7 peptide completely inhibited binding of caspase-9-p12 to BIR3. The affinity of the linker peptide and the Smac-N7 peptide towards BIR3 were comparable (linker, $K_d \approx 0.55 \pm 0.15 \mu\text{M}$; Smac-N7, $K_d \approx 0.81 \pm 0.18 \mu\text{M}$). Taken together, the above data indicate that Smac competes with caspase-9 for binding to the same pocket on the surface of XIAP, which could explain the ability of Smac to promote the catalytic activity of caspase-9 in the presence of XIAP. Consistent with this, mutation of a critical residue (E314), which is essential for binding to the Smac N-terminus and inhibition of caspase-9^{2,11}, abrogated binding of both Smac and caspase-9-p12 to BIR3 (Fig. 4b).

If the chemically synthesized linker peptide and processed caspase-9 bind to the same pocket on the surface of the BIR3 domain of XIAP, then we expect that the caspase-9 linker peptide should mimic the ability of Smac to promote caspase activation in S100 extracts in the presence of XIAP. To test this hypothesis, we examined the ability of the caspase-9 linker peptide or a peptide containing only the first five residues of the caspase-9-p12 to promote cytochrome

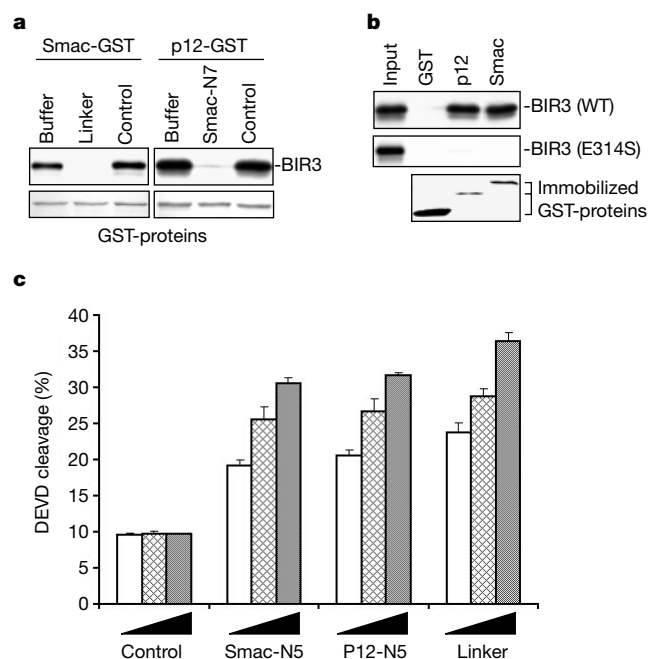


Figure 4 Binding of caspase-9-p12 or Smac to the BIR3 domain is mutually exclusive. **a**, Autoradiographic analysis of the interactions of ³⁵S-labelled BIR3 with Smac-GST or p12-GST proteins in the absence (buffer) or presence of 200 μM linker peptide (Smac-GST, left panel) or Smac-N7 peptide (p12-GST, right panel) or non-specific peptide (control). **b**, Autoradiographic analysis of the interactions of ³⁵S-labelled WT BIR3 or E314S mutant BIR3 with GST, p12-GST, or Smac-GST proteins. **c**, 293T S100 extracts were mixed with purified XIAP (10 nM) and then stimulated with cytochrome c plus dATP in the presence of increasing amounts of a non-specific peptide (control), Smac-N5, p12-N5 or linker peptides (25, 100, 500 μM). The reactions were carried out in the presence of DEVD-AMC as a substrate. Control peptide, MKSDFYFQK; Smac-N5 peptide, AVPIA; p12-N5 peptide, ATPFQ; linker peptide, ATPFQEGLRTFDQLD.

me-*c*-dependent activation of caspase-3 in S100 extracts containing XIAP (Fig. 4c). The activity of caspase-3 in the S100 extracts was measured using the peptide substrate DEVD-AMC. Both the linker and the p12-N5 peptides were as effective in promoting caspase-3 activation as the Smac-N5 or N7 (not shown) peptides in the XIAP-containing extracts. These results confirm that the linker peptide can compete with caspase-9 for binding to BIR3 and could function as an inhibitor of XIAP. During apoptosis, caspase-9 is further processed at Asp 330 by the activated caspase-3^{10,12}. On the basis of the above observations, processing at Asp 330 not only relieves the inhibition of caspase-9 by XIAP, but also releases the linker peptide into the cytoplasm, which could bind to XIAP and neutralize its inhibitory activity. Thus from a physiological aspect, the release of the linker peptide from caspase-9 during apoptosis constitutes a positive feedback loop in the potentiation of the caspase cascade and apoptosis.

Our observations reveal a mechanism for the activation and inhibition of caspase-9. Unlike other caspases, proteolytic processing of caspase-9 serves as a mechanism for inhibition, rather than

activation. In the absence of proteolytic processing, XIAP is unable to interact with procaspase-9 or inhibit its enzymatic activity. Upon apoptotic stimuli, procaspase-9 undergoes auto-catalytic processing in the context of a holoenzyme that contains Apaf-1 and cytochrome *c*, in which the apoptosome serves as the allosteric regulator of the caspase-9 activity^{10,12}.

In the crystal structure of a Smac–XIAP complex, the N-terminal tetra-peptide of Smac binds a surface groove on the BIR3 domain of XIAP (Fig. 5a), making a network of hydrogen-bond interactions and extensive van der Waals contacts¹³. The side chain of the first residue Ala fits in a conserved hydrophobic pocket in the surface groove of XIAP, which is formed in part by Trp 310 (Fig. 5a). The Ala main-chain groups hydrogen-bond to surrounding XIAP residues, including a pair of charge-stabilized hydrogen bonds to Glu 314. The high sequence similarity between the N-terminal sequences of the p12 subunit of caspase-9 and Smac predicts an identical mode of interaction with the BIR3 domain of XIAP (Fig. 5a). This is indeed supported by our experimental observations (Figs 3 and 4). The first residue Ala of the tetra-peptide is partially embedded in a pocket and uses its fully exposed amino group to hydrogen-bond to Glu 314, explaining why procaspase-9 must be proteolytically processed before it can bind XIAP (Fig. 2). In agreement with this prediction, mutation of Trp 310 or Glu 314 resulted in abrogation or significant reduction of interactions with both Smac and caspase-9 (Fig. 4, and data not shown).

The physical binding of the N-terminus of the caspase-9 p12 subunit to the BIR3 domain of XIAP constitutes an indispensable step in the inhibition of caspase-9. The close proximity of the N-terminus of the p12 subunit and the catalytic residue of caspase-9 suggests that XIAP may hinder the entry of the substrate to the active site (Fig. 5b). This proposed model is further supported by the observation that mutation of His 343 in XIAP-BIR3 results in complete loss of inhibition to the enzymatic activity of caspase-9², but not binding to caspase-9-p12 (data not shown), suggesting that His 343 directly binds the active site of caspase-9. Thus, although binding of the tetra-peptide of caspase-9 by XIAP appears to be a major contributor in their mutual interaction, other weaker interactions between caspase-9 and XIAP also contribute to the inhibition of caspase-9. Furthermore, the IAP proteins are likely to dimerize in solution⁷, which could further block substrate entry.

If Smac peptide interacts with the BIR3 domain of XIAP in the same manner as does caspase-9, how can Smac be more effective in relieving the inhibition of caspase-9? First, in addition to the tetra-peptide binding, Smac uses an extensive second interface to interact with the BIR3 domain of XIAP, involving over 2,000 Å² burial surface area¹³. This additional interaction may tip the balance in favour of the Smac–XIAP complex. Second, Smac also binds tightly to the BIR2 domain of XIAP⁷, which could facilitate the Smac–BIR3 interactions. Third, cleavage of caspase-9 after Asp 330 releases the linker peptide, which further helps to remove the inhibition of caspase-9 by XIAP. Finally, in apoptotic cells, the amount of Smac released from the mitochondria could be in excess.

The activation of procaspase-9 represents a critical step in the mitochondria-initiated apoptotic pathways. However, XIAP is unable to bind and inhibit procaspase-9 but binds and inhibits the proteolytically processed mature caspase-9. More strikingly, the mature caspase-9 uses the same conserved tetra-peptide to interact with XIAP as the mature form of Smac. These conserved interactions lead to opposing effects in caspase-9 activity and consequently apoptosis. □

Methods

cDNA cloning and expression of recombinant proteins

All caspase-9, caspase-3 and their mutant counterparts were overexpressed in *Escherichia coli* strain BL21(DE3) as C-terminally 6-histidine-tagged proteins using pET-21c or pET-28a vector (Novagen). Full length XIAP and its BIR3-RING (residues 243–497) or BIR1-BIR2 (residues 1–242) domain were overexpressed in *E. coli* strain DH5α as N-terminally

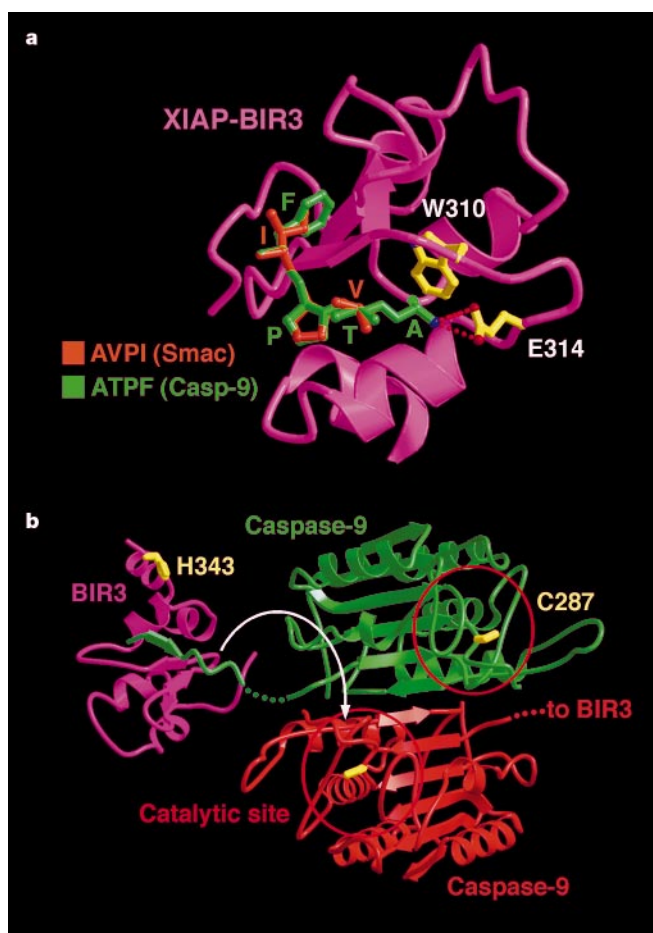


Figure 5 A proposed model of caspase-9 inhibition by XIAP. **a**, Conserved binding of BIR3 by caspase-9 and by Smac. XIAP-BIR3 is shown in pink. The N-terminal tetra-peptides from Smac (AVPI) and the p12 subunit of the human caspase-9 (ATPF) are shown in orange and green, respectively. Two critical residues on the BIR3 domain, W310 and E314, are highlighted in yellow. The N-terminal tetra-peptide (AVPY) from the p12 subunit of the rat or mouse caspase-9 more closely resembles the Smac peptide (AVPI).

b, A proposed model of caspase-9 inhibition by XIAP. The dimer of mature caspase-9 (based on the atomic coordinates of caspase-3, PDB code 1DD1) is represented by green and orange whereas the BIR3 domain of XIAP is shown in pink. The approximate location of the catalytic residue on caspase-9, C287, is highlighted in yellow. The catalytic site is identified by a red circle. H343, which is implicated in binding the catalytic site of caspase-9 is also shown in yellow.

GST-tagged proteins using a pGEX 4T vector (Pharmacia). Full-length XIAP, BIR3 and mutant BIR3 were *in vitro* translated in the presence of ^{35}S -methionine in reticulocyte lysates using MYC-pcDNA3 vector. Smac, caspase-9-p12, p10 and linker peptide were overexpressed in *E. coli* strain BL21(DE3) as C-terminally GST-tagged proteins using a pET-28-GST vector. Apaf-1L was expressed in Sf-9 insect cells and purified to homogeneity as previously described¹⁴.

Reconstitution and assay of the activity of the caspase-9–Apaf-1 holoenzyme complexes

Recombinant WT and mutant caspase-9 proteins (20 nM) were incubated in the presence or absence of recombinant Apaf-1 (20 nM) in buffer A⁶. The reaction mixtures were stimulated with cytochrome *c* (5 ng μL^{-1}) and dATP (1 mM) and incubated for 0–60 min at 30 °C with purified procaspase-3 C163A, and the processing of procaspase-3 was analysed by western blot analysis with anti-caspase-3 antibody (Fig. 1c). The holoenzyme complexes were also reconstituted as above in caspase-9-depleted S100 extracts from Apaf-1-deficient mouse embryonic fibroblasts (Fig. 1d). The DEVD-AMC cleaving activity of the reconstituted extracts were measured over a time of 0–120 min by luminescence spectrometry and represented in arbitrary spectrometric units. In some experiments the WT and the unprocessed triple-mutant (E306/D315/330A) caspase-9 proteins (specific activity ~10 fluorogenic units per s per ng) were reconstituted with Apaf-1 and cytochrome *c* plus dATP in the presence of increasing amounts of XIAP, and the cleavage of ^{35}S -labelled procaspase-3 by the complexes was analysed by SDS-PAGE and autoradiography (Fig. 2a).

Gel-filtration analysis of the caspase-9–Apaf-1 holoenzyme complexes

These procedures were performed as described before¹⁴. Caspase-9–Apaf-1 holoenzyme complexes containing WT or mutant caspase-9 were reconstituted in the presence of cytochrome *c* (5 ng μL^{-1}) and dATP (1 mM) in oligomerization buffer I (25 mM HEPES (pH 7.4), 50 mM NaCl, 10 mM KCl, 5 mM MgCl₂, 100 $\mu\text{g mL}^{-1}$ BSA, 5% glycerol, and 0.1 mM DTT) for 1 h at 30 °C. After incubation, the reaction mixtures were diluted with oligomerization buffer I and aliquots of each sample (100 μL) were loaded onto Superose-6 FPLC column. Equal volumes of the column fractions (50 μL) were separated by SDS-PAGE and immunoblotted with anti-Apaf-1 and anti caspase-9. In some experiments, the holoenzyme complexes were reconstituted in the presence of XIAP (Fig. 2b).

Far-western blotting

Affinity purified proteins were subjected to SDS-PAGE and were transferred to nitrocellulose membrane using standard western blotting techniques. The proteins on the membrane were denatured in a buffer (10 mM sodium phosphate pH 7.4, 150 mM sodium chloride, 5 mM magnesium chloride and 1 mM DTT) containing 6 M guanidine-HCl. The proteins were renatured by gradual reduction of guanidine-HCl to 0.3 M. The membrane was blocked overnight in the same buffer containing 5% non-fat dry milk and then incubated with ^{35}S -labelled *in vitro* translated XIAP in the same buffer with 1% non-fat dry milk. The membrane was washed at least three times and then exposed to X-ray film.

In vitro interaction assays

All *in vitro* interactions were performed as described⁶.

Determination of IC₅₀

WT caspase-3 or caspase-9 mutant proteins (10 pM) were incubated with purified BIR3 (0.5–800 nM) or BIR1-BIR2 proteins (0.1–80 nM) at 37 °C in the presence of DEVD-AMC (100 μM) for 30 min and the substrate cleavage was measured by luminescence spectrometry. The IC₅₀ values were then calculated using Graphpad Prism V2.0 (Graphpad Prism Software, San Diego, CA).

DEVD cleavage assays

Caspase-3 enzymatic assays with the tetrapeptide substrate DEVD-AMC were performed as described⁶.

Computer modelling

On the basis of the crystal structure of a Smac–BIR3 complex¹³, the N-terminal tetrapeptide of Smac was replaced by that from the p12 subunit of human caspase-9. Then we performed limited energy minimization on the complex between BIR3 and the tetrapeptide from the p12 subunit of the human caspase-9.

Received 4 December 2000; accepted 1 February 2001.

- Deveraux, Q. L. *et al.* Cleavage of human inhibitor of apoptosis protein XIAP results in fragments with distinct specificities for caspases. *EMBO J.* **18**, 5242–5251 (1999).
- Sun, C. *et al.* NMR structure and mutagenesis of the third BIR domain of the inhibitor of apoptosis protein XIAP. *J. Biol. Chem.* **275**, 33777–33781 (2000).
- Datta, R. *et al.* XIAP regulates DNA damage-induced apoptosis downstream of caspase-9 cleavage. *J. Biol. Chem.* **275**, 31733–31738 (2000).
- Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
- Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome *c*-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
- Srinivasula, S. M. *et al.* Molecular determinants of the caspase-promoting activity of Smac/DIABLO and its role in the death receptor pathway. *J. Biol. Chem.* **275**, 36152–36157 (2000).

- Chai, J. *et al.* Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature* **406**, 855–862 (2000).
- Abrams, J. M. An emerging blueprint for apoptosis in *Drosophila*. *Trends Cell. Biol.* **9**, 435–440 (1999).
- Chen, P., Nordstrom, W., Gish, B. & Abrams, J. M. Grim, a novel cell death gene in *Drosophila*. *Genes Dev.* **10**, 1773–1782 (1996).
- Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell* **1**, 949–957 (1998).
- Liu, Z. *et al.* Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **408**, 1004–1008 (2000).
- Rodriguez, J. & Lazebnik, Y. Caspase-9 and APAF-1 form an active holoenzyme. *Genes Dev.* **13**, 3179–3184 (1999).
- Wu, G. *et al.* Structural basis of IAP recognition by Smac/DIABLO. *Nature* **408**, 1008–1012 (2000).
- Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R. & Alnemri, E. S. Cytochrome *c* and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. *J. Biol. Chem.* **274**, 17941–17945 (1999).

Acknowledgements

We thank R. Penn for help with calculating IC₅₀ values. This work was supported by the NIH (E.S.A., P.D.R., Y.S.). S.M.S. is a special fellow of the Leukemia and Lymphoma Society; Y.S. is a Searle scholar and a Rita Allen scholar.

Correspondence and requests for materials should be addressed to E.S.A. (e-mail: E_Alnemri@lac.jci.tju.edu).

Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins

Monika Lachner*, Dónal O'Carroll*, Stephen Rea, Karl Mechtler & Thomas Jenuwein

Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, Dr Bohrgasse 7, A-1030 Vienna, Austria

* These authors contributed equally to this work

Distinct modifications of histone amino termini, such as acetylation, phosphorylation and methylation, have been proposed to underlie a chromatin-based regulatory mechanism^{1,2} that modulates the accessibility of genetic information. In addition to histone modifications that facilitate gene activity, it is of similar importance to restrict inappropriate gene expression^{3,4} if cellular and developmental programmes are to proceed unperturbed. Here we show that mammalian methyltransferases that selectively methylate histone H3 on lysine 9 (Suv39h HMTases)⁵ generate a binding site for HP1 proteins—a family of heterochromatic adaptor molecules^{6,7} implicated in both gene silencing and supra-nucleosomal chromatin structure. High-affinity *in vitro* recognition of a methylated histone H3 peptide by HP1 requires a functional chromo domain; thus, the HP1 chromo domain is a specific interaction motif for the methyl epitope on lysine 9 of histone H3. *In vivo*, heterochromatin association of HP1 proteins is lost in *Suv39h* double-null primary mouse fibroblasts but is restored after the re-introduction of a catalytically active SUV39H1 HMTase. Our data define a molecular mechanism through which the SUV39H–HP1 methylation system can contribute to the propagation of heterochromatic subdomains in native chromatin.

The 'histone code' hypothesis^{1,2} predicts that distinct modifications of the histone N termini can regulate interaction affinities for chromatin-associated proteins, similar to the binding of transcription factors to specific DNA sequences. For example, the acetylation of histone H4 tails induces an increased affinity for bromo-domain proteins⁸, thereby recruiting histone acetyltransferases (HATs) to nucleosomal regions of transcriptional activity. The human (SUV39H1) or murine (Suv39h1) histone

methyltransferases (HMTases) and their associated HP1 proteins⁹ share the evolutionarily highly conserved chromo domain^{10,11}, which has been implicated in directing euchromatic or heterochromatic localizations¹². On the basis of these intriguing parallels, we reasoned that the SUV39H1-dependent methylation of lysine 9 in H3 (ref. 5) may generate a specific binding surface for either the HMTase itself or for the associated HP1 protein(s).

To investigate directly whether SUV39H1 or HP1 proteins could display affinity for the H3 N terminus, we developed an *in vitro* 'pull-down assay' by covalently coupling unmodified, Lys-9-dimethylated (K9-dimeth) and Ser-10-phosphorylated (S10-phos) H3 N-terminal peptides (residues 1–20) to an affinity matrix (see Methods). As a control, we also prepared a matrix presenting an unmodified H4 N-terminal peptide (residues 1–20). Triple Myc-tagged complementary DNAs of human SUV39H1 and murine HP1 α , HP1 β and HP1 γ were *in vitro* translated, and recombinant proteins were incubated with a 1,000-fold or greater excess (see Methods) of the different covalently attached histone N-terminal peptides. After extensive washing, bound proteins were visualized by protein blot analysis with anti-Myc antibodies. To compare binding specificities, we also included the transcription factor Pax5 (ref. 13) and the murine Polycomb group protein mPC1 (ref. 14) in this analysis (Fig. 1a, b). The *Drosophila* Polycomb¹⁵ and HP1 (ref. 16) proteins have been shown to interact with histones *in vitro*.

Whereas (Myc)₃-Pax5 did not display binding to any of the matrices, (Myc)₃-mPC1 showed nonspecific associations. Similarly, although binding of (Myc)₃-SUV39H1 was slightly enhanced with the K9-dimeth peptide, it was also retained with all of the other N termini. In contrast, the three (Myc)₃-HP1 proteins only reacted with the K9-dimeth peptide (Fig. 1c, lower panels) but did not display associations with unmodified H3, or the S10-phos or H4 peptides.

HP1 proteins are characterized by the evolutionarily conserved chromo domain^{10,11} and by a related carboxy-terminal shadow

domain that directs intra- and intermolecular interactions^{6,7}. To examine the domain requirements for association with the H3 Lys 9 N terminus, we introduced point mutations (V23M or V26R) in the first β -sheet of the HP1 β chromo domain that would interfere with a predicted hydrophobic binding pocket¹². We also generated truncated (Myc)₃-HP1 β proteins that comprise only the chromo (residues 4–96) or the shadow (residues 110–185) domain (Fig. 2a). Quantitative pull-down assays with full-length HP1 β indicated that saturated binding occurred with a 500- to 10,000-fold excess of the H3 K9-dimeth peptide. By contrast, both chromo point mutants displayed impaired affinity and the shadow domain showed no interaction at all (Fig. 2b). Notably, the isolated HP1 β chromo domain can recognize the H3 Lys 9 methylation mark, although its binding requires a 10- to 20-fold increase in the concentration of the H3 K9-dimeth peptide as compared with full-length HP1 β .

To validate these findings, we next examined the relative ratios at which free peptide would compete away full-length HP1 β or the isolated HP1 β chromo domain from the affinity matrix. We bound about 20 ng of (Myc)₃-HP1 β or (Myc)₃-chromo to the H3 K9-dimeth matrix and then incubated with increasing concentrations (10-fold increments) of either unmodified or K9-dimeth H3 peptides. Whereas the unmodified H3 peptide was inactive over the entire concentration range, a 10,000-fold excess of the K9-dimeth H3 peptide was required to dissociate (Myc)₃-HP1 β (Fig. 2c). By contrast, the affinity of (Myc)₃-chromo was significantly weaker; this domain had already eluted at a 1,000-fold excess of the K9-dimeth H3 peptide. Together with the above results, these data indicate that the HP1 β chromo domain interacts with the H3 Lys9 methylation mark *in vitro*; however, they further suggest that full-length HP1 β or an HP1 β dimer¹⁷ is necessary for high-affinity binding. The latter interpretation is particularly intriguing because it would allow many interactions to occur between Lys-9-dimethylated H3 N termini and dimerized HP1 molecules, thereby stabilizing their binding.

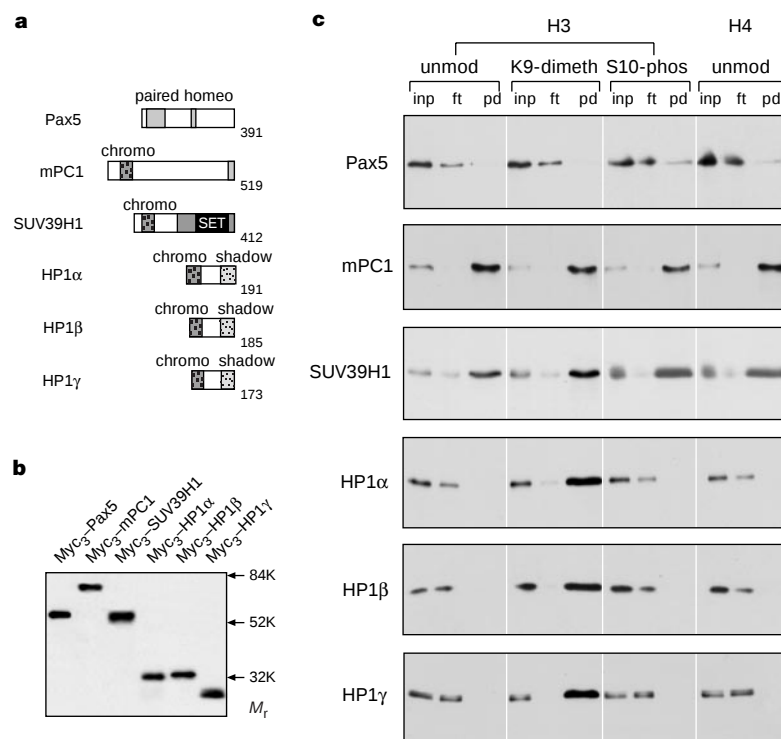


Figure 1 The H3 Lys 9 methyl epitope activates a specific binding site for HP1 proteins *in vitro*. **a**, Diagram of the domain structures of murine Pax5, murine mPC1, human SUV39H1 and murine HP1 α , HP1 β and HP1 γ . **b**, Protein blot of *in vitro* translated and pre-calibrated (Myc)₃-tagged proteins detected by anti-Myc antibodies. **c**, *In vitro* pull-

down assays with the indicated (Myc)₃-tagged proteins on affinity matrices that present either unmodified (unmod), Lys-9-dimethylated (K9-dimeth) and Ser10-phosphorylated (S10-phos) H3 N-terminal peptides, or an unmodified H4 N terminus. 'inp', 10% of IVT proteins used for binding; 'ft', 10% of flow-through; 'pd', pulled-down (bound proteins).

To show that H3 Lys 9 methylation is also the crucial determinant for HP1 localization in native chromatin, we next analysed the nuclear distribution of HP1 proteins in wild-type versus mutant primary mouse embryonic fibroblasts (PMEFs) in which the two murine *Suv39h* genes had been disrupted⁵. Mammalian HP1 proteins are enriched at heterochromatic foci^{18,19}, which can readily be visualized in mouse cells by counterstaining of (A+T)-rich repeat sequences with 4',6-diamidino-2-phenylindole (DAPI). Heterochromatic association of murine HP1 proteins was unaltered only if single *Suv39h1* or *Suv39h2* gene disruptions were introduced into PMEFs (data not shown; for wild-type localization see below). By contrast, indirect immunofluorescence for HP1 β or HP1 α epitopes in interphase chromatin of *Suv39h* double-null PMEFs indicated that both proteins are dispersed from heterochromatin. These results are consistent with complementary analyses of *clr4* mutants in *Schizosaccharomyces pombe*²⁰.

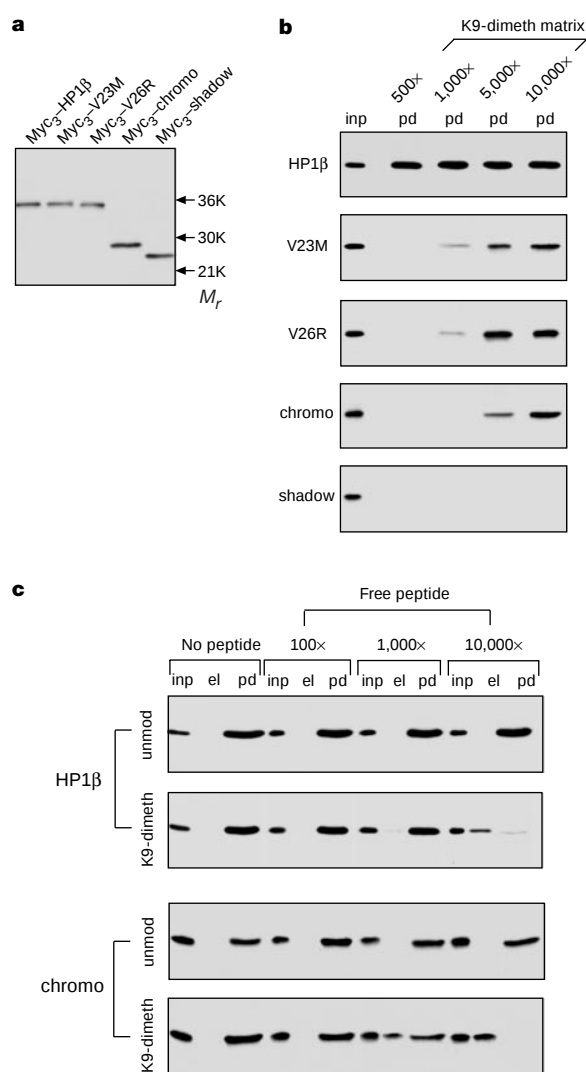


Figure 2 Mutant analysis of HP1 β and quantitation of *in vitro* binding to the H3 K9-dimeth peptide matrix. **a**, Protein blot of *in vitro* translated and pre-calibrated (Myc)₃-tagged HP1 β mutant proteins detected by anti-Myc antibodies. **b**, Quantitative pull-down assays of (Myc)₃-tagged HP1 β mutants with the H3 K9-dimeth affinity matrix. Binding was analysed with ~20 ng of IVT proteins and increasing concentrations of H3 K9-dimeth N termini (indicated as fold excess). **c**, Quantitative elution of pre-bound (Myc)₃-HP1 β and (Myc)₃-chromo with increasing concentrations (indicated as fold excess) of unmodified or K9-dimeth H3 N-terminal peptides. 'inp', 10% of pre-bound proteins; 'el', 20% of first wash; 'pd', pulled down (proteins retained on matrix after final washes).

To determine more definitively that HP1 localization is dependent on Suv39h-mediated HMTase activity and not on association with Suv39h proteins^{9,21}, we re-introduced an active human (Myc)₃-SUV39H1 HMTase or a catalytically inactive (Myc)₃-SUV39H1 mutant (H324L)⁵ into wild-type and *Suv39h* double-null PMEFs by retroviral infection. Expression of a linked green fluorescent protein (GFP) indicated 80–90% infection efficiency (see Methods; data not shown). Protein blot analysis with nuclear extracts from infected cells confirmed a comparable abundance of ectopic proteins (Fig. 3a), and catalytic activity or inactivity (of the H324L mutant) was shown by HMTase assays after immunoprecipitation with anti-Myc antibodies (Fig. 3b).

We then examined nuclear localization of ectopic SUV39H1 proteins in wild-type PMEFs by indirect immunofluorescence with anti-Myc antibodies. Whereas mock-infected cells were Myc negative, both (Myc)₃-SUV39H1 and the (Myc)₃-SUV1(H324L) mutant are enriched at heterochromatic foci, in a staining that largely overlaps with colocalized HP1 β and HP1 α (Fig. 4, right panels). This result indicates that heterochromatin association of SUV39H1 in a wild-type background can be uncoupled from its intrinsic HMTase activity, and is consistent with our previous structure–function analysis that defined a heterochromatin targeting and HP1 β interaction domain at the N terminus of SUV39H1 (ref. 21).

Finally, we colocalized ectopic SUV39H1 and endogenous HP1 β and HP1 α proteins in infected *Suv39h* double-null PMEFs. In contrast to wild-type cells, the (Myc)₃-SUV1(H324L) mutant only displayed diffuse nuclear staining, similar to the dispersed distribution of endogenous HP1 proteins (Fig. 4, left panel). Notably, the catalytically active (Myc)₃-SUV39H1 HMTase was enriched at heterochromatin and could induce relocation of

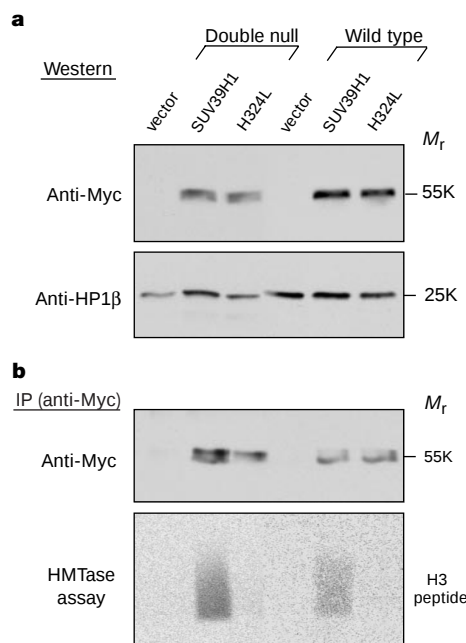


Figure 3 Characterization of transduced SUV39H1 HMTase activity in PMEFs. **a**, Protein blot analysis with 30 μ g of nuclear extracts from wild-type and *Suv39h* double-null mouse primary embryonic fibroblasts (PMEFs) that had been mock-infected (vector) or productively infected with retroviruses carrying an active ('SUV39H1', (Myc)₃-SUV39H1) or an inactive ('H324L', (Myc)₃-SUV1(H324L)) HMTase. As a loading control and to demonstrate similar abundance of endogenous HP1 proteins, the blot was cut and also probed with antibodies specific for HP1 β . **b**, HMTase assay with an N-terminal H3 peptide and transduced (Myc)₃-SUV39H1 enzymes that were enriched by immunoprecipitation (IP) from 500 μ g of nuclear PMEF extracts with anti-Myc antibodies.

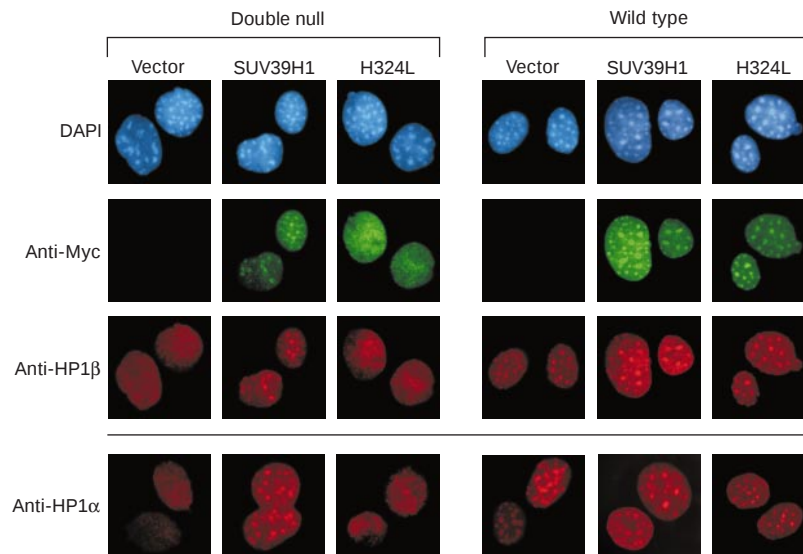


Figure 4 Rescue of heterochromatic localization of HP1 proteins by an active SUV39H1 HMTase. Double-label indirect immunofluorescence of wild-type (right panel) and *Suv39h* double-null (left panel) PMEFs that had been infected with (Myc)₃-SUV39H1-transducing retroviruses as in Fig. 3. Expression analysis of a linked GFP protein indicated 80–90% infection efficiency (data not shown). Mouse interphase chromatin was co-stained with

anti-Myc (green) and anti-HP1β antibodies (red). DNA was counterstained with DAPI (blue). In additional samples, PMEF interphase chromatin was co-stained with anti-Myc (not shown) and anti-HP1α antibodies (red). Contrast of images was adjusted with Adobe Photoshop 4.0.

HP1β and HP1α to these heterochromatic foci (Fig. 4, middle row, left panel). Together, these results strongly suggest that the Suv39h-dependent methylation of Lys 9 in H3 (ref. 3) transduces the signal to recruit the SUV39H–HP1 complex to heterochromatic regions in mammalian chromatin.

In contrast to gene activity controlled by sequence-specific transcription factors, epigenetic mechanisms for variegating gene expression have been difficult to define at a molecular level, because it remained unclear how they would discriminate target loci for activation or repression^{22–25}. The Suv39h HMTases selectively methylate H3 at Lys 9, and their activity is impaired by pre-existing modifications in the H3 N terminus⁵. Thus, conditions that entail the presentation and incorporation of unmodified H3 tails, for example, decaying transcriptional activity during developmental programmes²⁶ or replication-coupled histone deposition^{27,28} at transcriptionally non-permissive regions, are likely to favour the establishment of the Suv39h-dependent H3 Lys 9 methyl epitope. Moreover, because Lys 9 in H3 can either be methylated or acetylated^{1,29}, competitive modification of this position could provide a molecular ‘switch’ for the induction of euchromatic or heterochromatic subdomains, if the activity or abundance of the respective enzymes is altered. For example, forced overexpression of SUV39H1 induces ectopic heterochromatin by redistributing endogenous HP1β²¹. On the basis of our data, we propose that the SUV39H–HP1 methylation system contributes to higher order chromatin organization by propagating a heterochromatin-specific affinity in the histone H3 N terminus. □

Methods

In vitro translation of (Myc)₃-tagged proteins

A (Myc)₃–H6 epitope including a unique *NotI* cloning site was inserted into the polylinker of pKW2T, which allows *in vitro* transcription by the SP6 RNA polymerase. The plasmid for parental (Myc)₃-SUV39H1 (residues 3–412) has been described⁹. Additional (Myc)₃-tagged proteins were generated by transferring *NotI*/*EcoRI* polymerase chain reaction (PCR) amplicons into the pKW2T derivative, encoding in-frame fusions for Pax5 (residues 4–391)¹³, mPC1 (residues 3–519)¹⁴, HP1α (residues 5–181), HP1β (residues 5–185), HP1γ (residues 5–173) and the HP1β chromo (residues 4–96) or HP1β shadow (residues 110–185) domain^{7,11,17}. The V23M and V26R point mutations were directly generated in the (Myc)₃-HP1β construct by double PCR mutagenesis. All plasmids were confirmed by sequencing, which detected the presence of a point mutation (G8A) in

mPC1 and of another point mutation (D65A) in HP1γ.

Plasmid DNA (1 μg) was transcribed and translated (TNT-kit; Promega) in a volume of 50 μl, and *in vitro* translated (IVT) proteins were calibrated by immunoblotting with anti-Myc antibodies visualized by peroxidase staining using enhanced chemiluminescence (ECL; Amersham).

In vitro pull-down assays with peptide matrices

Affinity matrices were generated in 400 μl of HEPES pH 7.9 by covalently coupling, for 2 h at room temperature, 1.4 mg of histone peptides (~600 nmol) through a C-terminal cysteine to 5 mg of activated Poros beads (Applied Biosystems), representing ~300 nmol of active maleimide groups through the bifunctional crosslinker sulfo-GMBS (Pierce). Saturation of the matrices was determined by the 412-nm absorbance of free thiol-groups using 5,5'-dithio-bis (2-nitrobenzoic acid) (Sigma). Matrices can be stored at 4 °C.

The unmodified H3 N-terminal (residues 1–20), K9-dimeth and S10-phos H3 peptides have been described⁵. In addition, we also prepared an affinity matrix with the N terminus of histone H4 (SGRGKGGKGLGKGGAKRHRK-Cys).

In vitro pull-down assays were modified on the basis of described protocols³⁰ with a ~1,000-fold excess of N termini over IVT proteins. For one *in vitro* reaction, 10 μl of the peptide matrices (≤ 1 nmol of covalently attached peptide) were washed three times in binding buffer (20 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100) before incubation with 2–10 μl of IVT proteins (~1 pmol or 20 ng of recombinant protein) in a volume of 170 μl for 60 min at room temperature. Reactions were vigorously washed five times in 50 mM HEPES pH 7.5, 0.5 M NaCl, 1 mM EDTA, 1% NP-40, and bound proteins were eluted in SDS sample buffer, separated by 10 or 12% SDS-PAGE and visualized by immunoblotting with anti-Myc antibodies.

For the elution analysis, eight aliquots of 20 ng IVT proteins were bound for 1 h at room temperature under saturating conditions to 10 μl ((Myc)₃-HP1β) or 100 μl ((Myc)₃-chromo) of the K9-dimeth matrix and washed twice. Loaded matrices were then incubated for an additional 1 h at room temperature in 150 μl of binding buffer containing increasing concentrations of free peptides, washed five times and analysed as above.

Analysis and rescue of PMEFs

PMEFs were derived from embryonic day E12.5 *Suv39h* double-null fetuses obtained after intercrossing *Suv39h1*^{1–1}/*Suv39h2*^{1–1} compound mutant mice (D. O’C. *et al.*, manuscript in preparation). As controls, PMEFs were prepared from wild-type fetuses of the same genetic background. Co-staining of Triton X-100 extracted PMEF interphase chromatin was performed with rat monoclonal anti-HP1β¹⁸ and mouse monoclonal anti-Myc (9E10) antibodies, or with mouse monoclonal anti-HP1α (Piere Chambon, Illkirch) and rabbit polyclonal anti-Myc (9E10) (Lukas Huber, IMP) antibodies, followed by counterstaining of DNA with 4',6'-diamidino-2-phenylindole (DAPI; Sigma) as described²¹.

For the rescue experiment, the H324L mutation⁵ was introduced into full-length (Myc)₃-SUV39H1, and (Myc)₃-SUV39H1 and (Myc)₃-SUV1(H324L) *SalI*/*EcoRI* DNA fragments were transferred into the polylinker of MSCV–MIGR1 (a derivative of murine stem-cell virus plasmid; Clontech) that contains a linked humanized IRES–GFP protein instead of PGK–*neo*. SUV1 plasmids and empty MSCV–MIGR1 vector were transiently transfected into the ecotropic φNX-eco helper-free packaging cell line (G. Nolan, Stanford). Passage 3 wild-type and *Suv39h* double-null PMEFs (3 × 10⁵ cells per 10-cm dish) were infected twice with φNX virus S/N, and cells were grown for 3 d before analysis

of GFP expression, indirect immunofluorescence, ectopic protein abundance and HMTase assay³.

Received 9 November; accepted 8 December 2000.

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
2. Turner, B. M. Histone acetylation and an epigenetic code. *BioEssays* **22**, 836–845 (2000).
3. Paro, R. & Harte, P. J. in *Epigenetic Mechanisms of Gene Regulation* (eds Russo, V. A. E., Martienssen, R. A. & Riggs, A. D.) 507–528 (CSHL, New York, 1996).
4. Pirrotta, V. Polycomb: the genome, PcG, trxG, and chromatin silencing. *Cell* **93**, 333–336 (1998).
5. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
6. Wallrath, L. L. Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**, 147–153 (1998).
7. Jones, D. A., Cowell, I. G. & Singh, P. B. Mammalian chromodomain proteins: their role in genome organisation and expression. *BioEssays* **22**, 124–127 (2000).
8. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAF_{II}250 double bromodomain module. *Science*, **288**, 1422–1425 (2000).
9. Aagaard, L. *et al.* Functional mammalian homologues of the *Drosophila* PEV modifier *Su(var)3–9* encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
10. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl Acad. Sci. USA* **88**, 263–267 (1991).
11. Ball, L. J. *et al.* Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO J.* **16**, 2473–2481 (1997).
12. Platano, J. S., Harnett, T. & Eisenberg, J. C. Functional analysis of the chromodomain of HP-1. *EMBO J.* **14**, 3977–3986 (1995).
13. Adams, B. *et al.* Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.* **6**, 1589–1607 (1992).
14. Pierce, J. J., Singh, P. B. & Gaunt, S. J. The mouse has a polycomb-like chromobox gene. *Development* **114**, 921–929 (1992).
15. Breiling, A., Bonte, E., Ferrari, S., Becker, P. B. & Paro, R. The *Drosophila* Polycomb protein interacts with nucleosomal core particles *in vitro* via its repression domain. *Mol. Cell. Biol.* **19**, 8451–8460 (1999).
16. Zhao, T., Heyduk, T., Allis, C. D. & Eisenberg, J. C. Heterochromatin protein 1 (HP1) binds to nucleosomes and DNA *in vitro*. *J. Biol. Chem.* **275**, 28332–28338 (2000).
17. Brasher, S. V. *et al.* The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *EMBO J.* **19**, 1587–1597 (2000).
18. Wreggett, K. A., Hill, E., James, P. S., Hutchings, G. W. & Singh, P. B. A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenet. Cell Genet.* **66**, 99–103 (1994).
19. Minc, E., Allory, Y., Worman, H. J., Courvalin, J.-C. & Buendia, B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. *Chromosoma* **108**, 220–234 (1999).
20. Ekwall, K. *et al.* Mutations in the fission yeast silencing factors *clr4*⁺ and *rik1*⁺ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *J. Cell Sci.* **109**, 2637–2648 (1996).
21. Melcher, M. *et al.* Structure–function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation and mitotic progression. *Mol. Cell. Biol.* **20**, 3728–3741 (2000).
22. Grosfeld, F. Activation by locus control regions? *Curr. Opin. Genet. Dev.* **9**, 152–157 (1999).
23. Festenstein, R. *et al.* Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal context-dependent manner. *Nature Genet.* **23**, 457–461 (1999).
24. Nielsen, A. L. *et al.* Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J.* **18**, 6385–6395 (1999).
25. Ryan, R. F. *et al.* KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* **19**, 4366–4378 (1999).
26. Brown, K. E. *et al.* Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* **91**, 845–854 (1997).
27. Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
28. Taddei, A., Roche, D., Sibarita, J.-B., Turner, B. M. & Almouzni, G. Duplication and maintenance of heterochromatin domains. *J. Cell. Biol.* **147**, 1153–1166 (1999).
29. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **103**, 263–271 (2000).
30. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. Histone H3 and H4 N-termini interact with Sir3 and Sir4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* **80**, 583–592 (1995).

Acknowledgements

We would like to thank M. Busslinger for the *Pax5* cDNA; P. B. Singh for the *HP1β* (M31) and *mPC1* (M33) cDNAs, and the *HP1β* antibodies; A. Verreault for the murine *HP1α* and *HP1γ* cDNAs; P. Chambon for *HP1α* antibodies; L. Huber for rabbit anti-Myc (9E10) antibodies; and Y. Zou for the MSCV–MIGR1 retroviral vectors. We acknowledge I. Gorny for help with peptide synthesis and M. Doyle for contributing to the *Suv39h2* knock-out. We are grateful to D. Allis for discussions and to M. Busslinger for comments and critical reading of the manuscript. Research in T.J.'s laboratory is supported by the IMP through Boehringer Ingelheim, the Austrian Research Promotion Fund and the Vienna Economy Promotion Fund.

Correspondence and requests for materials should be addressed to T.J. (e-mail: jenuwein@nt.imp.univie.ac.at).

Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain

Andrew J. Bannister^{*†}, Philip Zegerman^{*†}, Janet F. Partridge[‡], Eric A. Miska^{*}, Jean O. Thomas[§], Robin C. Allshire[‡] & Tony Kouzarides^{*}

^{*} Wellcome/CRC Institute and Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK
[‡] MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
[§] Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

[†] These authors contributed equally to this work

Heterochromatin protein 1 (HP1) is localized at heterochromatin sites where it mediates gene silencing^{1,2}. The chromo domain of HP1 is necessary for both targeting and transcriptional repression^{3,4}. In the fission yeast *Schizosaccharomyces pombe*, the correct localization of Swi6 (the HP1 equivalent) depends on Clr4, a homologue of the mammalian SUV39H1 histone methylase^{5,6}. Both Clr4 and SUV39H1 methylate specifically lysine 9 of histone H3 (ref. 6). Here we show that HP1 can bind with high affinity to histone H3 methylated at lysine 9 but not at lysine 4. The chromo domain of HP1 is identified as its methyl-lysine-binding domain. A point mutation in the chromo domain, which destroys the gene silencing activity of HP1 in *Drosophila*³, abolishes methyl-lysine-binding activity. Genetic and biochemical analysis in *S. pombe* shows that the methylase activity of Clr4 is necessary for the correct localization of Swi6 at centromeric heterochromatin and for gene silencing. These results provide a stepwise model for the formation of a transcriptionally silent heterochromatin: SUV39H1 places a 'methyl marker' on histone H3, which is then recognized by HP1 through its chromo domain. This model may also explain the stable inheritance of the heterochromatic state.

The histone code hypothesis predicts the existence of domains that specifically recognize modified histone residues⁷. One such domain is the bromodomain, which allows transcription factors such as P/CAF and TAF_{II}250 to recognize histone tails only when they are acetylated at lysine residues^{8,9}. To identify an equivalent domain that recognizes methyl-lysines, we examined a number of chromatin-associated factors for their ability to recognize specifically a histone H3 peptide (residues 1–16) tri-methylated at lysines 4 and 9.

Figure 1b shows that one protein, HP1, was able to bind to the methylated peptide (red line), but not to the unmethylated peptide (data not shown), using surface plasmon resonance (SPR) analysis. Other chromatin-associated proteins, such as Mi2, MeCP2 and P/CAF did not bind in this assay (data not shown). Of the two conserved domains in HP1 (Fig. 1a), the chromo domain was able to bind (Fig. 1b, green line) whereas the chromo-shadow domain was not (blue line). Figure 1c shows that the on-rate for the chromo domain (red line) is not affected by the addition of unmethylated histone H3 peptide (green line), but binding is completely abolished in the presence of the methylated H3 peptide (blue line). The dissociation constant (*K_d*) of the HP1 chromo domain for the methylated H3 peptide (methyl-lysine) is about 70 nM, as determined by SPR (data not shown). The affinity of the HP1 chromo domain for methyl-lysine compares favourably with the affinity of the TAF_{II}250 and P/CAF bromodomains for acetyl-lysine⁸, although a more definitive comparison will be possible with ITC analysis of the HP1/methyl-lysine interaction. Our analysis of other chromo domain proteins such as polycomb (M33), Mi2 and

SUV39H1 has not identified another chromo domain with methyl-lysine-binding ability (data not shown).

The specificity of the HP1 chromo domain for methyl-lysine can also be shown using a 'pull-down' assay with H3 peptide linked to beads (Fig. 2a). Full-length wild-type HP1 that is bacterially expressed, labelled with 35 S-labelled methionine (input), bound the methylated H3 peptide but not the unmethylated peptide. A large proportion of the input protein is bound, which is consistent with the low K_d measured by SPR analysis.

Mutations of conserved residues in the HP1 chromo domain (V23M and KW41/42AA) abolish recognition of methylated H3, whereas mutation of a non-conserved residue (L68P) has little effect on binding (Fig. 2a). The chromo domain of Mi2 does not bind the methylated H3 peptide in this assay. These results indicate the high affinity and specificity of the HP1 chromo domain structure in the recognition of methylated lysines. One of the chromo domain mutants (V23M) disrupts the gene silencing functions of *Drosophila* HP1 (ref. 3), suggesting a link between the ability of HP1 to repress transcription and its ability to bind methyl-lysine of histone H3.

We next tested whether the HP1 chromo domain had a preference

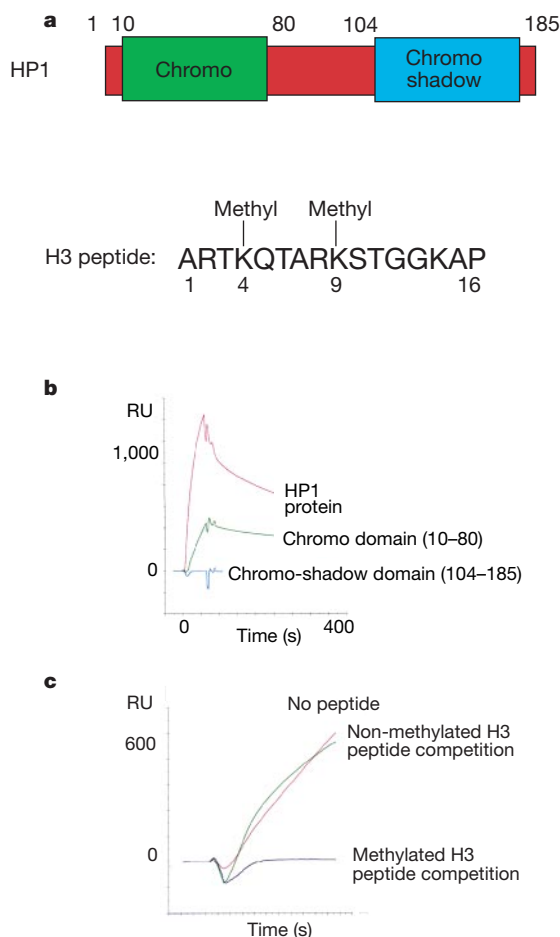


Figure 1 HP1 chromo domain binds methylated H3. **a**, Schematic representation of HP1 and histone H3 peptide used in this study. Peptides were either unmodified, tri-methylated at both Lys 4 and Lys 9, or tri-methylated at Lys 4 or Lys 9. **b**, Equimolar amounts of his-HP1 full length (red line), his-chromo domain (green line) and the his-chromo-shadow domain (blue line) were tested for binding to the Lys 4 and Lys 9 tri-methylated histone H3 peptide using a BIAcore J instrument. The y-axis units are resonance units (RU), which are proportional to the mass of the protein binding to the peptide on the surface of the chip. **c**, His-HP1 was tested as described for **b** except that it was pre-incubated with either no peptide (red line), unmethylated peptide (green line) or Lys 4 and Lys 9 tri-methylated peptide (blue line).

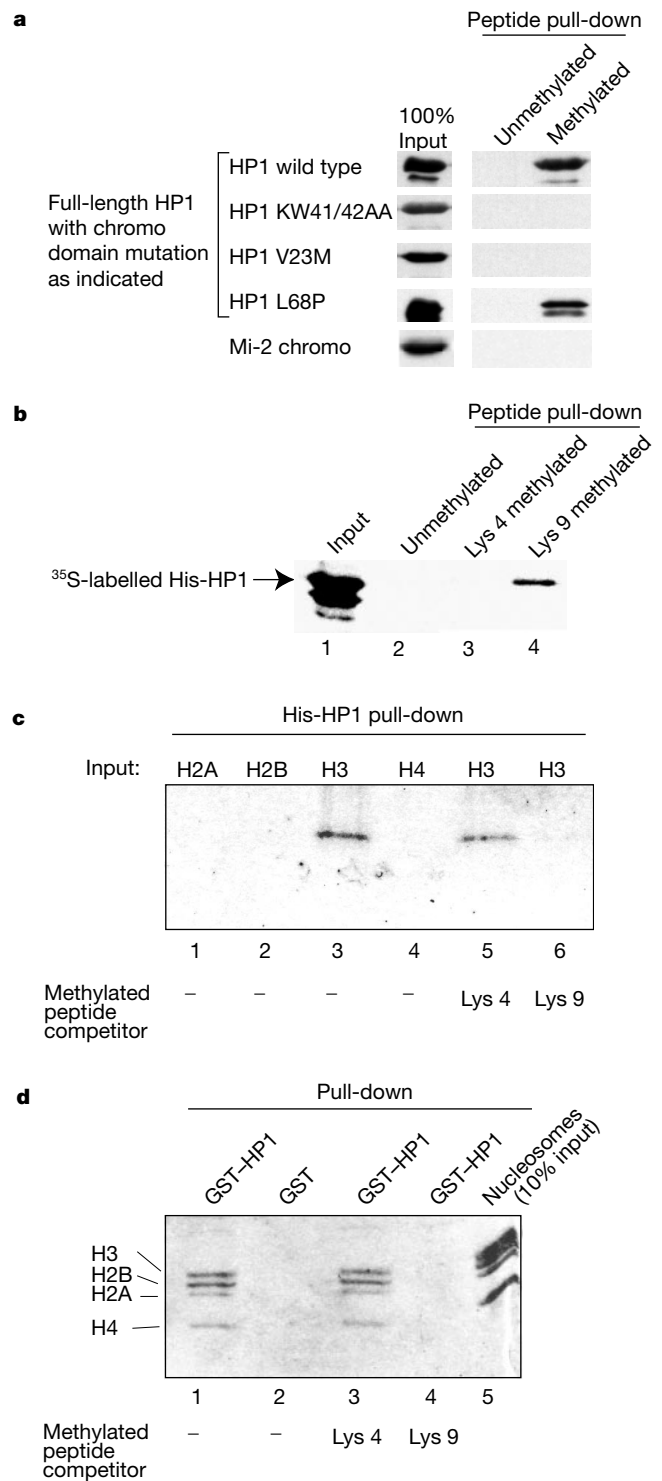


Figure 2 HP1 chromo domain specifically binds methylated Lys 9 of histone H3.

a, Recombinant, radiolabelled proteins were tested for binding to either unmodified or Lys 4 and Lys 9 tri-methylated H3 peptide (see Fig. 1a) immobilized on beads. **b**, His-HP1 full length was tested for binding to unmethylated peptide (lane 2), Lys 4 methylated peptide (lane 3) or Lys 9 methylated peptide (lane 4). Lane 1 shows 100% input. **c**, Each core histone (5 μ g) was assessed for binding to his-HP1 in a pull-down experiment without competitor peptide (lanes 1–4), or in the presence of histone H3 peptide tri-methylated at Lys 4 (lane 5) or Lys 9 (lane 6). After SDS–PAGE, bound histone was visualized by Coomassie staining. **d**, Mono-nucleosomes (200 μ g) were assessed for binding to GST/HP1 or GST without competitor peptide (lanes 1 and 2), or in the presence of histone H3 peptide tri-methylated at Lys 4 (lane 3) or Lys 9 (lane 4). Ten per cent input is shown (lane 5). After SDS–PAGE, nucleosome-binding was determined by Coomassie staining of nucleosomal histones.

for a particular methyl-lysine within histone H3. H3 peptides methylated at either Lys 4 or at Lys 9 were individually tested for HP1 binding (Fig. 2b). HP1 shows clear selectivity for methylated Lys 9 as it does not bind to methylated Lys 4. HP1 recognized dimethylated Lys 9 as efficiently as tri-methylated Lys 9, and all mammalian forms of HP1 (α , β and γ) show the same selectivity for methylated Lys 9 (data not shown).

To test whether HP1 could recognize specifically full-length methylated histone H3, we added Ni^{2+} -agarose-bound his-HP1 to individually purified histones (Roche) and performed a pull-down assay. Figure 2c shows that his-HP1 only binds to histone H3 (lane 3) and not to histones H2A, H2B or H4 (lanes 1, 2 and 4). The H3 recognized by HP1 is methylated at Lys 9 because the HP1/H3 interaction can be disrupted by excess methylated Lys 9 H3 peptide (lane 6), but not by methylated Lys 4 H3 peptide (lane 5).

The ability of HP1 to recognize methylated histone H3 caused us to investigate whether this recognition allows HP1 to associate with methylated mono-nucleosomes. Glutathione S-transferase (GST)/HP1 can bind a proportion of purified nucleosomes from chicken erythrocytes (see Methods), whereas GST does not (Fig. 2d, compare lanes 1, 2 and 5). The association between HP1 and nucleosomes is not disrupted by the addition of excess methylated Lys 4 H3 peptide (lane 3) but is completely disrupted by the addition of excess methylated Lys 9 peptide (lane 4). Thus, HP1 can recognize mono-nucleosomes, at least partly, through its recognition of Lys 9-methylated H3 tails.

We next assessed whether HP1 was tethered to cellular chromatin through its specific recognition of methylated histone H3. We prepared nuclei from human U2OS cells such that they were depleted of most soluble protein. They were then challenged with methylated or unmethylated H3 peptide, and after the insoluble (chromatin) fraction was pelleted, the supernatant was western blotted for HP1 using an HP1 antibody that recognizes

all mammalian HP1 forms¹⁰. Figure 3a shows the proportion of HP1 in total nuclear fraction (lane 1), compared with the soluble fraction after incubation (lane 2). Challenging with excess methylated H3 peptide increases the proportion of HP1 α , β and γ in the soluble fraction (lane 4), whereas competition with the unmethylated H3 peptide had no effect (lane 3). Thus, methylated H3 peptide can displace HP1 from the chromatin fraction of U2OS nuclei *in vitro*.

We next asked whether competition with the methylated H3 peptide could disturb the localization of HP1 in U2OS cells. Cell fixation and permeabilization conditions were established whereby a peptide could be delivered into cells, and then immunofluorescence was used to detect HP1. Figure 3b shows the nuclear localization of HP1 when no peptide was added; this did not change when an unmethylated H3 peptide was added (Fig. 3c). However, addition of the methylated H3 peptide resulted in displacement of HP1 from the nucleus, where it is otherwise held by association with chromatin (Fig. 3d), and accumulation in the cytoplasm. Similar results were obtained in a distinct cell line, CH1 (data not shown).

The fission yeast *S. pombe* has an HP1 homologue Swi6, which contains a chromo domain that is closely related to those of HP1 family members¹. Figure 4a shows that the Swi6 chromo domain is able to bind H3 peptide methylated at Lys 9, whereas its chromo-shadow domain has no methyl-binding activity. *S. pombe* also contains an SUV39H1 homologue, Ctr4, which possesses methylase activity that is specific for histone H3 Lys 9 (ref. 6). To investigate whether the methylase activity of Ctr4 is necessary for the correct localization of Swi6 in yeast, we examined a point mutant (G341D) in the SET domain of Ctr4, which has lost the ability to methylate histone H3 (Fig. 4b). When the localization of Swi6 is monitored in Ctr4-G341D mutant yeast cells, its distribution is different from that in wild-type cells. Swi6 is localized to two to four foci in wild-type

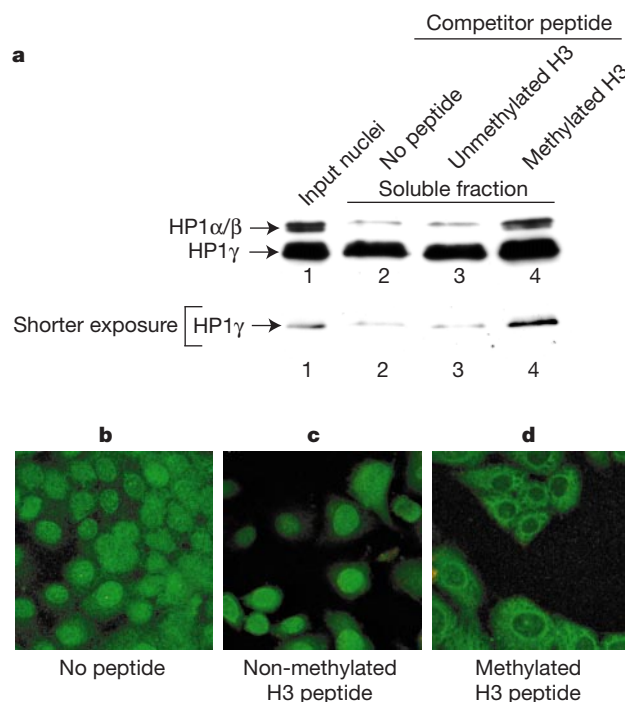


Figure 3 HP1 chromatin localization is dependent on binding to methylated K9 of H3. **a**, Purified and permeabilized U2OS cell nuclei were challenged with no peptide (lane 2), unmethylated peptide (lane 3) or the Lys 4 and Lys 9 methylated peptide (lane 4). After incubation, supernatant was western blotted for HP1 using an antibody specific for HP1 α , β and γ (long exposure, top; shorter exposure, bottom). Lane 1

shows 50% input. **b**, HP1 immunofluorescence in U2OS cells without competitor peptide. **c**, **d**, Immunofluorescence as in **b**, except that unmethylated H3 peptide or the Lys 4 and Lys 9 methylated peptide was added, respectively. The peptide did not affect the specificity of the antibody (data not shown).

cells, corresponding to silent chromatin domains at the nuclear periphery¹¹, whereas the *Clr4*-G341D strain shows loss of localization from the nuclear periphery and accumulation of more diffuse staining over the nucleolus (Fig. 4c). The *Clr4*-G341D strain shows a very similar pattern of Swi6 distribution to that of the *Clr4Δ* strain lacking *Clr4* (ref. 5). The specificity of the Swi6 antibody is demonstrated using a *Swi6Δ* strain (Fig. 4c).

To confirm that Swi6 association with silent chromatin is dependent on SET domain function, we performed chromatin immunoprecipitation with anti-Swi6 antibodies (Fig. 4d). Centromeric chromatin, known to be associated with Swi6, was enriched relative to the control (euchromatin) locus in the immunoprecipitated sample compared with the total extract¹². In both the *clr4*-G341D and *clr4Δ* strains, enrichment of this centromeric sequence was lost. Thus, a mutation in the *Clr4* SET domain that abolishes histone H3 methylase activity disrupts recruitment of Swi6 to silent chromatin. Loss of association of Swi6 with centromeres should result in expression of a normally silent marker gene embedded in centromeric chromatin. Figure 4e shows that this is indeed the case. On indicator plates, wild-type strains silence the centromeric *ade6*⁺ marker, which results in red, repressed colonies¹³; however, in strains lacking *Clr4* (Δ) or strains defective in *Clr4* methylase activity (G341D), this *ade6*⁺ gene is clearly expressed, resulting in the formation of white colonies. Collectively these results show that methylase activity of *Clr4* is required for the recruitment of Swi6 to silenced centromeric heterochromatin and for transcriptional silencing.

The experiments described here fit well with existing data showing that SUVAR39H1 and HP1 colocalize at heterochromatic sites and interact biochemically⁵. We show that HP1 recognizes

histone H3 only when Lys 9 has been methylated by SUVAR39H1. In *S. pombe*, Lys 9 methylase activity is required for transcriptional repression and HP1 localization. Thus, these data make a direct correlation between methylated, HP1-bound histones and transcriptionally silent heterochromatin.

The fact that HP1 is associated with the enzyme that methylates and 'marks' histones for HP1 binding suggests a self-maintenance model for how HP1 spreads over chromatin to form heterochromatically repressed regions (Fig. 5). An extension of this model could also explain how silenced heterochromatin could be passed on during DNA replication. The SUVAR39H1 methylase, bound to methylated histones via HP1, could direct the methylation of newly deposited histones. This self-maintaining mechanism would then

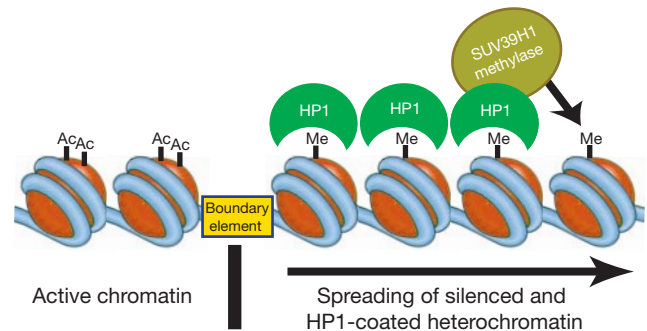


Figure 5 A model of heterochromatic self-maintenance by the SUVAR39H1/HP1 complex.

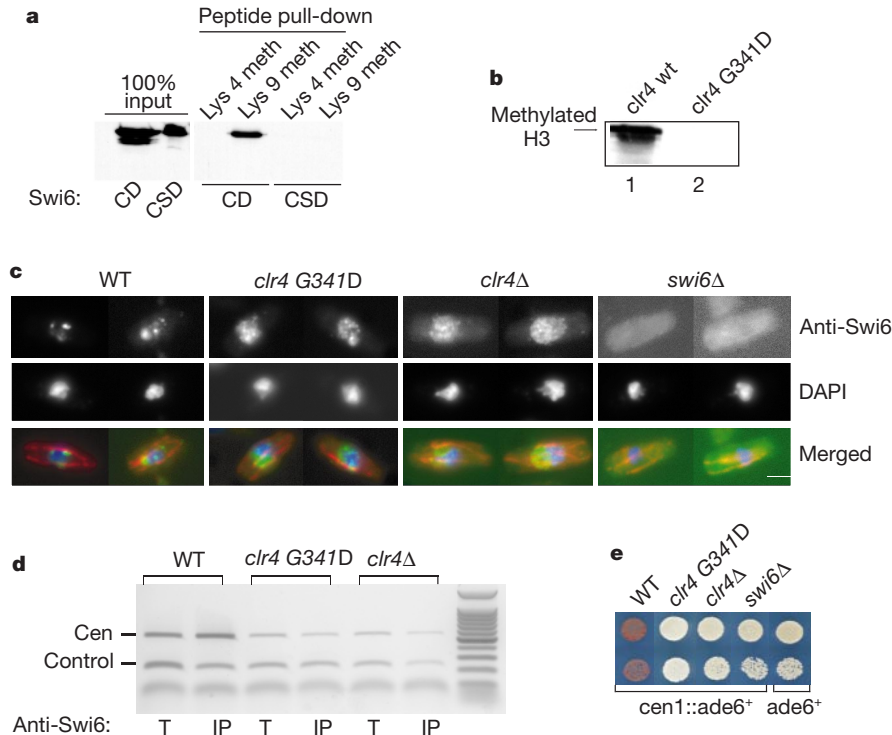


Figure 4 Swi6 localization is dependent on the methylase activity of the *Clr4* SET domain. **a**, Swi6 chromo domain (CD) or chromo-shadow domain (CSD) were tested for binding to H3 peptide methylated at either Lys 4, or Lys 9. **b**, Recombinant *Clr4* SET-domain protein (or a mutant version, G341D) were tested for histone methylase activity as described⁶. **c**, Swi6 immunolocalization in strains bearing wild-type *clr4* (WT), a complete deletion of *clr4* (*clr4Δ*), or a missense mutation in the *Clr4* SET domain (*clr4* G341D). Swi6 immunolocalization (top), DAPI staining of chromosomal DNA (middle) and a merged

image of α -tubulin staining (red), DAPI (blue) and Swi6 (green) (bottom) are shown. Scale bar, 5 μ M. **d**, Chromatin immunoprecipitation using Swi6 antibodies. T, total extract input; IP, immunoprecipitated sample. Multiplex PCR assessed enrichment of centromeric (cen) sequences over an euchromatic control in the different strain backgrounds. **e**, Strains bearing mutations in *clr4* and *swi6* were tested for their ability to transcriptionally silence the centromeric *ade6*⁺ marker on media with limiting adenine (red colonies, repressed; white colonies, expressed). An *ade6*⁺ strain (white) provides a control.

allow the spread of HP1 on new histones, re-establishing the silent heterochromatic state in the next generation. □

Methods

Peptides and biophysical techniques (SPR)

H3 peptide: NH₂-ARTKQTARKSTGGKAPGGC-COOH. We incorporated methyl-lysines where required. Peptides were immobilized on BIAcore CM5 chips, and interactions were measured in HBS-EP buffer (10 mM HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA and 0.005% v/v surfactant P20) using a BIAcore J Instrument.

Pull-down assays

Recombinant proteins were expressed in and purified from *Escherichia coli* as described¹⁴. We produced radiolabelled protein in bacteria by growing them in the presence of ³⁵S-labelled methionine. GST fusions of the Swi6 chromo domain (residues 80–133) and chromo-shadow domain (residues 261–328) have been described¹⁵. Mouse full-length HP1 was expressed either as a GST fusion protein or as a his-tagged protein. Mouse HP1 KW41/42AA and V23M mutants were expressed as GST fusions and L68P as a his-tagged protein. The Mi-2 chromo domain was expressed as a his-tagged protein. Where appropriate, we eluted recombinant proteins from beads and dialysed them against HBS-EP.

Peptides, linked to SulfoLink Coupling Gel (Pierce), were used in pull-down assays with eluted GST or his-tagged fusion proteins in HBS-EP. Pull-down assays with histones were done in 600 mM NaCl HBS-EP, whereas nucleosomes were analysed in 400 mM NaCl HBS-EP.

Preparation of nuclei and peptide challenge

U2OS nuclei were purified¹⁶ and permeabilized¹⁷ as described, except that 0.25% Triton X-100 was used. Permeabilized nuclei were diluted into PBS and incubated with or without peptide for 2 h on ice. We pelleted nuclei and analysed the supernatant by western blotting.

Immunofluorescence of mammalian cells

U2OS cells were fixed for 2 min in ice-cold methanol (containing 10 µg ml⁻¹ peptide, where used) and then blocked for 15 min in 3% bovine serum albumin, 0.6% Triton-X-100 in PBS (containing 10 µg ml⁻¹ peptide, where used). We performed staining using anti-HP1 antibodies¹⁰ (1:1000). Antibody incubations contained 20 µg ml⁻¹ peptide (where used).

Schizosaccharomyces pombe strains and media

All strains had a centromeric insertion of the *ade6⁺* marker (*otr1R-Sph1::ade6⁺*), and carried a genomic mutation in *ade6* (*ade6-210*), with the exception of the *ade6⁺* strain used as a control in the silencing assay. *swi6Δ::hist⁺* was described previously¹⁸, and *clr4Δ::ura4⁺* was generated by replacement of the *clr4⁺* open reading frame with the *ura4⁺* marker. The *clr4* G341D allele was isolated as a mutant that alleviated telomeric silencing¹⁹. We assessed centromeric silencing as described²⁰.

Protein methyltransferase assay

Wild-type and mutant (G341D) *clr4* SET-domain regions (residues 127–490) were cloned into pMAL (NEB) and expressed as maltose-binding protein fusions. Methyltransferase assays were done as described⁶.

Yeast immunostaining

We performed immunostaining as described⁵, except that we used Texas Red-conjugated anti-mouse, and FITC-conjugated anti-rabbit antibodies. Cells were simultaneously stained with mouse monoclonal antibodies against α-tubulin (TAT1) and rabbit anti-Swi6 antibodies.

Chromatin immunoprecipitation

Swi6 antibodies immunoprecipitated fixed chromatin as described¹². Immunoprecipi-

tated DNA was analysed by PCR using primers that detect the *imr/otr* junction of the outer repeats of the centromere and the euchromatic *flp1⁺* locus (control).

Preparation of mono-nucleosomes

We isolated chicken erythrocyte mono-nucleosomes as described²¹.

Received 8 November; accepted 8 December 2000.

1. Eissenberg, J. C. & Elgin, S. C. R. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **10**, 204–210 (2000).
2. Eissenberg, J. C. *et al.* Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **87**, 9923–9927 (1990).
3. Platero, J. S., Hartnett, T. & Eissenberg, J. C. Functional analysis of the chromo domain of HP1. *EMBO J.* **14**, 3977–3986 (1995).
4. Cavalli, G. & Paro, R. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. *Curr. Opin. Cell Biol.* **10**, 354–360 (1998).
5. Ekwall, K. *et al.* Mutations in the fission yeast silencing factors *clr4⁺* and *rik1⁺* disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *J. Cell Sci.* **109**, 2637–2648 (1996).
6. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
7. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
8. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
9. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tijan, R. Structure and function of a human TAF₂₅₀ double bromodomain module. *Science* **288**, 1422–1425 (2000).
10. Pak, D. T. S. *et al.* Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* **91**, 311–323 (1997).
11. Ekwall, K., Olsson, T., Turner, B. M., Cranston, G. & Allshire, R. C. Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell* **91**, 1021–1032 (1997).
12. Partridge, J. F., Borgstrom, B. & Allshire, R. C. Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* **14**, 783–791 (2000).
13. Ekwall, K. *et al.* The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269**, 1429–1431 (1995).
14. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
15. Cowieson, N. P., Partridge, J. F., Allshire, R. C. & McLaughlin, P. J. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Curr. Biol.* **10**, 517–525 (2000).
16. Krude, T. Mimosine arrests proliferating human cells before onset of DNA replication in a dose-dependent manner. *Exp. Cell Res.* **247**, 148–159 (1999).
17. Krude, T. Initiation of human DNA replication *in vitro* using nuclei from cells arrested at an initiation-competent state. *J. Biol. Chem.* **275**, 13699–13707 (2000).
18. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–233 (1995).
19. Nimmo, E. R., Pidoux, A. L., Perry, P. E. & Allshire, R. C. Defective meiosis in telomere-silencing mutants of *Schizosaccharomyces pombe*. *Nature* **392**, 825–828 (1998).
20. Allshire, R. C., Javerzat, J. P., Redhead, N. J. & Cranston, G. Position effect variegation at fission yeast centromeres. *Cell* **76**, 157–169 (1994).
21. Thomas, J. O. in *Chromatin, a Practical Approach* (ed. Gould, H. J.) 1–34 (Oxford Univ. Press, 1998).

Acknowledgements

We thank E. Laue and N. Murzina for his-HP1 clones, M33 and Mi2 clones, and for discussions on structure, and P. Chambon for GST-HP1 clones. We also thank D. Durocher for advice with SPR, T. Krude for guidance on nuclear preparations and M. Ruas for preparing some proteins. We are grateful to E. Andrews for help in the preparation of nucleosomes, and to E. Nimmo and P. Lord for the *clr4-G341D* allele. Peptides were synthesized by G. Bloomberg, Bristol University. This work was funded by a Cancer Research Campaign programme grant to T.K. and Medical Research Council core support to R.A. J.O.T. thanks the BBSRC for support.

Correspondence and requests for materials should be addressed to T.K. (e-mail: tk106@mole.bio.cam.ac.uk).

Scanning the spectrum

Programs, gadgets and instrumentation for spectroscopy and fluorimetry.



Four in-line: Varian's Peltier cell holder.

Peltier

Varian

www.varianinc.com

Multi-cell holder for fluorescence spectroscopy

This thermostatted, four-position cell holder is suitable for DNA thermal denaturation/renaturation, enzyme kinetics, membrane and protein studies with the Cary Eclipse fluorescence spectrophotometer. The Peltier cell holder features electromagnetic stirring and electronic heat pumps, known as peltiers, that improve the accuracy and precision of temperature control. Suggested applications include cell membrane analysis and signal transduction studies.

Reader Service No. 100

Lambda

PerkinElmer

www.perkinelmer.com

UV/vis spectrometers

Lambda 25, 35 and 45 double-beam spectrometers offer a choice of optical resolution and stray-light performance. All models operate under full UV WinLab software, now upgraded for kinetics (UV KinLab) and biochemical analysis (UV BioLab). Noise and stray light are nominal, maximizing performance in both high- (up to 3 A) and low-absorbance (0.003 A) measurements, making the instruments suitable for studying precious samples such as dsDNA that are available only in small quantities.

Reader Service No. 101

Raman spectroscopy

Glen Spectra

www.jyhoriba.co.uk

Filter tips

A new brochure from Omega Optical, available from Glen Spectra, describes the Omega

range of filters for Raman application, including laser line interference filters to clean up the excitation source, alpha epsilon edge filters, XR Raman sets, rejection band filters, Raman notch filters, tunable narrow band filters and Raman microscope laser reflectors. e-mails@jyhoriba.co.uk for a copy.

Reader Service No. 102

UV/visible spectrophotometers

Amersham Pharmacia Biotech

www.apbiotech.com

Compact and curvaceous

Designed for nucleic-acid and protein quantification, this range of spectrophotometers can perform a range of functions from simple colorimetric assays to stringent research. Some models are GLP- and Pharmacopoeia-compli-



Bright idea: Amersham Pharmacia's new range.

ant and all come with built-in application software. The brochure "Ultra visible. UV/visible spectrophotometry" is available from distributors and from the company website.

Reader Service No. 103

2102 Series

Unico

www.unico1.com

Space-saving spectrophotometer

With a narrow bandpass of 2 nm, the Model 2102 and 2102A UV/vis scanning spectrophotometers are designed for use in research, quality control and biochemical work. Advanced applications are controlled via Windows-based software, while simple functions can be performed using the instrument's control panel. To save bench space, the 2102 can accommodate a 15-inch monitor on top.

Reader Service No. 104

IL1715

International Light

www.intl-light.com

All-in-one radiometer/photometer

The IL1715 system eliminates the need for separate instruments and detectors to make visible and radiometric measurements. It includes the IL1700 instrument, SED033 detector, Y photopic filter, F flat response filter, W diffuser, optical bench post and carrying case, and provides direct readings in watts per cm², ft-candles and lux. Featuring automatic zeroing and automatic ranging during exposure integrations over an 8.5-decade range, the IL1715 has a photometric dynamic range of 5×10^{-4} to 1×10^6 lux and a radiometric dynamic range of 1×10^{-9} to 2×10^{-1} W cm⁻². Applications include a variety of flash and continuous wave light measurements.

Reader Service No. 105

Mini V

Shimadzu

www.shimadzu.com

High performance in a compact device

The Mini V visible spectrophotometer can be used to perform a range of quantitative methods from simple concentration measurements to sophisticated calibration curves. It comes with a spectrum mode that allows for full spectral data acquisition from 325 to 1,100 nm. Methods, results and raw data can be saved either on the instrument, on optional dedicated IC data cards or on a PC using UV Data Manager software. Print options include a thermal screen copy printer or a variety of PC printers. Accessories include multicell positioners, sippers and temperature control devices.

Reader Service No. 106

new on the market

Stingray

Bio-Rad

www.bio-rad.com

Imaging ATR spectrometer

The Stingray attenuated total reflection imaging spectrometer acquires all spectral information from all the pixels in the detector to provide both spectral and spatial resolution on samples that are not amenable to either transmission or reflection analysis. The technique includes imaging of both microscopic samples using an infrared focal plane array (FPA) mounted on an infrared microscope accessory, and macroscopic samples, where a conventional sample compartment is used in conjunction with the FPA.

Reader Service No. 107

Spectral ID 3

Galactic

www.galactic.com

Powerful search software

Spectral ID Version 3 is compatible with over 750,000 spectra in commercial libraries including Aldrich, Bio-Rad Sadtler, Chemical Concepts, Nicolet-Aldrich, NIST and Wiley. It can create and search libraries for many techniques including IR, Raman, MS, UV-vis, fluorescence and NIR. Complete GC-MS or GC-IR data sets can be searched and molecular structures displayed, edited and stored. The program includes algorithms for mass spectral searching and supports remote library management. Optional software can be used on Windows NT servers to give each user access to all the installed libraries.

Reader Service No. 108

ProxiPlate-384 F

Packard Instrument

www.packardinstrument.com

Shed more light on the matter

Suggested for use with Packard's homogeneous time-resolved fluorescence (HTRF) technology in the Discovery microplate analyser, the black ProxiPlate-384F is made from low-fluorescence resin and is suitable for fluorescence applications in which sterility is not required. It has a shallow well with a total volume of 30 μ l. Samples are positioned close to the detectors, producing enhanced light-signal output, sensitivity and effective miniaturization, with reduced reagent costs.

Reader Service No. 109

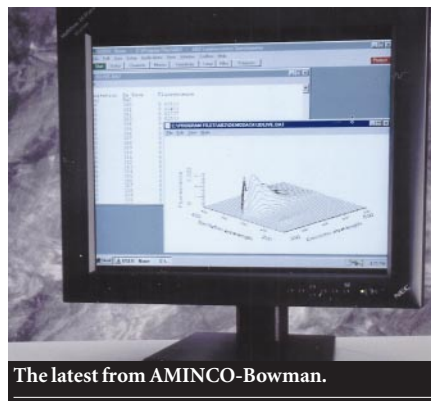
AB2 software

Spectronic Unicam

www.uvvis.com

In the frame

Windows applications software is now available for the AMINCO-Bowman Series 2 luminescence spectrometer. Users wishing to upgrade from OS/2 can transfer current data files directly to Windows 95/98. The new software allows users to create personalized setups, execute multiple programs simultaneously



ly and share data between programs. Graphics capabilities include three-dimensional representation with 'point-of-view' perspective options, complete control of axes, customized legends and a variety of display modes. Analysis options include smoothing, blank subtraction, normalization and curve fitting.

Reader Service No. 110

GRAMS/32 A1 6

Adept Scientific

www.adeptscience.co.uk

Get an assistant

Version 6 of the GRAMS/32 A1 spectroscopy data-handling program is compatible with Windows 2000 and offers enhanced data processing based on Microsoft ActiveX technology and an updated interface featuring an assistant to guide users through the software. This upgraded version can read data files from hundreds of analytical instruments including NMR, FT-IR, Raman, NIR, GC-MS, UV-vis, and fluorescence instrumentation.

Reader Service No. 111

Gene Spec III

Hitachi Instruments

www.hii.hitachi.com

1024-element photodiode array

Gene Spec III scans and quantifies DNA/RNA as well as protein samples. Specialized software measures absorbance, 260/280 nm ratio, concentration and purity of DNA/RNA. Other capabilities include wavelength scanning, time scans and quantitation. Samples and their results are measured and displayed in a complete spectrum from 190 nm to 110 nm in as little as eight seconds. Quartz cuvettes are available in 1, 2, 5, 10, 20 and 50 μ l sizes. Operations are controlled via a PC, and data and spectra can be displayed or stored in one of 20 channels and printed later in Excel.

Reader Service No. 112

FTIRsearch

Nicolet

www.ftirsearch.com

Online database

Logging on to this website allows users to search high-quality spectral databases using common web browsers such as Netscape Navigator or Internet Explorer. A new feature saves the

results of the last search conducted for retrieval at a later date. Search credits can be purchased online in Belgium, France, Germany, Italy, Japan, Luxembourg, the Netherlands and the United Kingdom, as well as in the United States and Canada using a purchase order in addition to the standard online credit card purchase. Library collections are indexed and available as PDF files for both existing users and non-account-holding visitors to the site.

Reader Service No. 113

TSQ Quantum

ThermoFinnigan

www.thermoquest.com

Hyped-up transmission

Measuring 55 $\times$ 56 $\times$ 79 cm, this mass spectrometer has a quadrupole rod structure described by the manufacturer as truly hyperbolic. A 90° collision cell with square quadrupole technology optimizes ion transmission and increases signal-to-noise ratio, resulting in high resolution and minimal matrix interference. The complementary Xcalibur software suite can be used for quantitative metabolite identification, drug discovery and proteomic applications.

Reader Service No. 114

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

BACs • YACs • PACs cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics

U.S. or Canada 800-533-4363

U.K. 0-800-89-1393

FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 6

DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING



10K-50K
Compounds
in Plates

- Next Day Delivery*
 - Competitively Priced
- *Some Restrictions Apply

Chembridge Corporation • 800.964.6143
858.451.7400 • 858.451.7401 fax • E-mail: chem@chembridge.com

www.chembridge.com

READER ENQUIRY NO. 60

**At a crossroads**

Sandia scientists expand work at disciplinary interfaces
p128

**Wanted**

Specialists to design and run new instrumentation
p129

Better microscopes will be instrumental in nanotechnology development

Customizing and combining existing tools may help advance the fledgling field of nanoscience, says Steve Bunk.

Ever since the 1980s, when IBM researchers unveiled the scanning tunnelling microscope (STM) and the atomic force microscope (AFM), instruments that explore or probe the surface of ultra-tiny samples have steadily proliferated. Physicists, chemists and molecular biologists working in nanoscience have had to familiarize themselves with microscopes that use thermal, magnetic, capacitive and electrochemical properties. But what next for this formidable family of scanning probe microscopes (SPMs)?

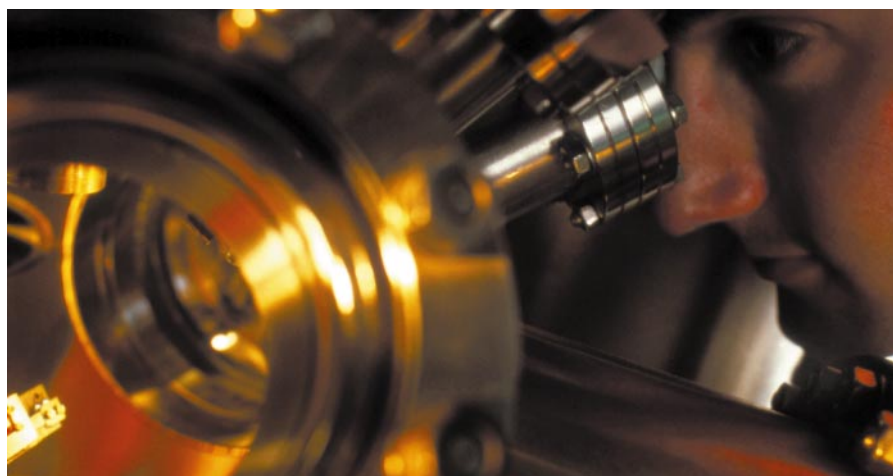
The short answer is that researchers at the leading edge of nanoscience expect yet more bells and whistles, whereas those with less specialized needs may benefit from a burgeoning of user-friendly SPMs. The long answer concerns the divide between those needs. "There is a distinction between technology and science," says Phaeton Avouris, manager of IBM's Nanoscale Science and Technology Group in New York. "In reality, there is no nanotechnology yet."

No product will emerge soon from research at the nano-level, Avouris says. All commercial applications are still on the microscale. He likens the probe or tip of an SPM to a finger: it can touch, push, drag and not much else. For the nanoscale, that is not enough. "In the future, we have to go to more complicated schemes, where we can manipulate in three dimensions, for example," he says.

The difficulty in progressing from experimentation to technology lies not only in an ability to build reproducible nanoscale structures, but in connecting a nanodevice to macro-controls. Until these problems are resolved, SPMs and other tools such as optical traps and microneedles, that can pin down or apply force to molecules and other tiny things, will remain the staples of nanoscience.

The current challenge does not lie in operating such instruments but in understanding what is observed with them. This could require familiarity with chemistry, surface science and even electronic structure.

But finding and developing people with multiple skills in physics, biology and chem-



Focus of interest: the scanning tunnelling microscope has helped to revolutionize nanoscience.

istry is a "hot-button issue," says Stanford University biophysics professor Steven Block. A leader in the study of biological motors, he uses custom-built instrumentation for detecting displacements at the nanometre level. He says nanoelectromechanical systems (NEMS) is a sexy term as yet without substance. The here and now is microelectromechanical systems (MEMS), which increasingly are attempting to emulate natural ways of functioning on the nanoscale. Most current research in the field involves movement of microscale materials, such as single cells, through nanoscale displacements. "In the near term, the most exciting developments will come from where biology meets MEMS, not where biology meets NEMS," he says.

Winning combinations

To help to advance that cause, the next generation of instruments will combine the functions of several devices. For example, capabilities in single-molecule fluorescence, and in a technique called fluorescence resonance energy transfer that can measure distances on the 2–8-nm scale, might be combined with optical trapping to generate new quantitative data. Objects then could be interrogated with light and grasped while molecular forces and phase changes

are measured, all at high resolution. Or the near-field scanning optical microscope (NSOM), which offers improved resolution over standard optical tools, might be combined with an AFM, to yield imaging and force measurement.

"There's a lot of interest now in developing 'smart handles,' a whole set of specific attachments used by biologists," Block says. For instance, nickel histidine, glutathione or antibody epitopes could be used on microscope tips as 'grabber' molecules. A problem with current SPMs is getting the tips to release the molecules that they grasp.

Such work with single molecules is a far cry from the so-called universal assembler, a molecular 'factory' that one day may achieve the goal of directed self-assembly by repeatedly constructing nanoscale devices. Block says that this goal encounters the most intractable problems of emergent nanotechnology: how to grab a single atom with certitude, how to make a bond with the grabber, how to keep the two from getting stuck together, and how to manipulate objects on a nanoscale using a nanoscale device.

Surmounting those obstacles will require clever computational modelling, applied mathematics, and thermodynamic simulations, says Rodney Ruoff, a Northwestern

careers and recruitment

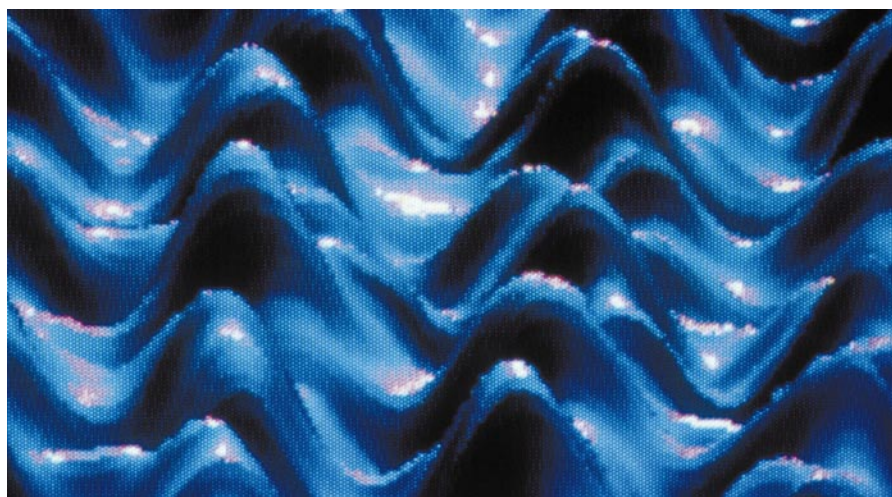
► University mechanical engineering professor. He foresees the development of a through-space method of controlling self-assembled nanodevices, or an on-board instruction set, to sever physical ties with the macro-world.

Ruoff has customized components of SPMs, such as a mechanical stage built inside an electron scanning microscope that enables the loading of multiwalled carbon nanotubes for observation. He is now working on a much smaller stage that will fit inside a tunnelling electron microscope, and is adamant that students should learn instrument building.

Future investments

Thomas Weber, director of the materials research division at the National Science Foundation (NSF) in Washington, estimates he has about \$14 million annually to fund instrumentation, not counting portions of grants to principal investigators that might go towards equipment. The NSF's major research instrumentation programme, which funds a broad range of science and engineering equipment for universities and is not under Weber's management, has a 2001 budget of about \$75 million.

The NSF is the lead organization in the new National Nanotechnology Initiative, which was allocated \$423 million this year. A 1999 interagency working group report that outlines goals in the development of investigative tools over the next 5–10 years lists instruments with sub-surface and three-dimensional imaging capabilities at nanometre-scale spatial resolution. The report also prioritizes developing nanomanipulators to build molecules or clusters and combine them into larger devices, and parallel probe arrays for full characterization



A fresh view: new equipment allows researchers to view items on the nanoscale.

of a nanostructure's properties.

Paul West, a physical chemist who has been developing microscopes since 1983, is now vice-president of marketing for ThermoMicroscopes of Sunnyvale, California, a leading supplier of SPMs. West says his team has been supporting development of a new instrument called the nanoManipulator for five years; University of North Carolina researchers had been working on it for about five years before that. An STM with a graphics display and a feedback mechanism that fits on a finger, it is now marketed by ThermoMicroscopes. West says there are multiple probe systems on the horizon, perhaps with touch-feedback interfaces on three to five fingers. But he cautions that what is achievable and what the market will bear can be quite different.

"We see more new ideas than probably any single business," he declares. Deciding

what can be successfully commercialized is a major challenge, West says. "A lot of the ideas we run across are great, but they would take a long time to develop."

He points to the concept of mass information storage using parallel AFM tips to read and write data by sensing and making indentations on a polymer surface. The idea has been around for almost 20 years, he says, but nothing is yet on the market. IBM's latest prototype of its Millipede project, revealed late last year, uses more than 1,000 AFM tips and has achieved data densities of 100 to 200 gigabits per square inch, several times higher than current state-of-the-art magnetic storage devices. But IBM does not expect to commercialize the product for three to five years.

"In general, the development of these technologies has gone much faster than scientists can absorb them," West says. The

Sandia scientists develop instruments at the crossroads



Under control: researchers at Sandia depositing thin films on wafers.

Standing in a hallway at Sandia National Laboratories in Albuquerque, New Mexico, physicist Alan Burns declares: "We work at interfaces." It is true in several senses. In the surface and interface sciences department, Burns and colleagues do microtechnology research that crosses into the nanoscale. They observe and move molecules, with a special interest in activity at another interface, the cellular membrane. And this research carries them to intersections of their scientific disciplines. Mostly physicists and chemists, the team is now in search of both junior and senior molecular biologists and biophysicists.

Burns looks at the wall, on which material is pinned, to help him describe a key invention, aptly called the interfacial force microscope

near-term will see dramatic improvements in the manufacturing of probes for all SPM applications.

A large share of the market for laser-assisted microscopic manipulation, including the registered LaserTweezers for optical trapping, and LaserScissors for microdissection and microsurgery, is commanded by Cell Robotics of Albuquerque, New Mexico. Clients are beginning to require customizations such as fluorescence capabilities, says Cell Robotics product manager Larry Keenan. Chief scientist Jerome Conia comments: "What we have seen lately is a change in these requirements, where scientists are more interested in subtle effects, and are not so much into power as resolution." That change, he adds, signals their deepening emphasis on the nano-world.

A handful of nanotechnology start-ups have appeared in recent years, following the 1997 debut of Zyvex in Richardson, Texas. Ralph Merkle, a nanotechnology theorist, left his job as a research scientist at Xerox PARC to become principal fellow at Zyvex in 1999. He says that the company, which has about 30 employees, is looking for people skilled in chemical assembly, scanning technology, MEMS design, computational chemistry, and computational modelling. "We'll be stirring all these people together in one pot."

Early this year, Zyvex announced a two-year development agreement with manufacturer Standard MEMS of Burlington, Massachusetts. Merkle says there is interest in building an instrument to make a microscale driver or mechanical device that could be detached from a wafer and reattached elsewhere.

You can never have enough new instruments, says IBM fellow Donald Eigler. He

recently completed building his second IBM microscope from the ground up, at the Almaden Research Center in San Jose, California. It is an ultrahigh-vacuum STM, to allow investigation of structures cooled to half a degree Kelvin by application of a magnetic field. Despite advances made in manufacturing, Eigler and others must still build their own machines to do certain types of research. "We made a count once in my family, and I think we have about 20 different kinds of saws," he muses. "The reason is, there's a lot of things to get done."

Steve Bunk is a science writer based in New Mexico.

Mass-spectrometry experience is in demand

Two companies with their headquarters in different countries dominated the race to build automated high-throughput gene sequencers. Now Europe-based Amersham Pharmacia Biotech and Applied Biosystems, based in the United States, are competing to build the best equipment to discern the identity, structure and function of proteins.

The competition is healthy for both the field and scientists seeking a job within it. Proteomics is still a relatively small-scale, slow affair, especially compared with high-throughput genome sequencing. Commercial proteomics equipment focuses on two tasks: separating proteins and identifying them. In order to boost sensitivity, through-

National Nanotechnology Initiative

♦ <http://www.nano.gov>

A practical guide to SPMs

♦ <http://www.thermomicromscopes.com>

Academic, government and industry nanoscience links

♦ <http://www.nanoindustries.com>

Nanotechnology

♦ <http://www.ioppublishing.com/Journals/na>

Nano Letters

♦ <http://pubs.acs.org/journals/nalefd/index.html>

Sandia National Laboratories

♦ <http://nano.sandia.gov>

Search tool for nanotechnology on the web

♦ <http://nanospot.org>



Information overload: many companies are now moving into protein analysis.

put and automation of proteomics equipment such as mass spectrometers, industry will need to draw on a diverse set of skills,



(IFM). Unlike the atomic force microscope (AFM), the Sandia instrument copes well with the adhesive forces of molecules touched by its tip. Those forces can cause the AFM probe to 'snap to contact' with the sample and release suddenly, perturbing the data. Instead of an AFM's cantilever, which bends because of the forces between tip and sample, the IFM has a sensor like a seesaw, that uses force-feedback to balance the torque applied to the tip.

Physicist Jack Houston invented the IFM about a decade ago at Sandia, but Burns says it took years to iron out the bugs. Although the technology has been licensed and commercialized, only a handful of institutions currently have them. "The AFM is a perfect example of an instrument that was a real curiosity at first, and now it's a workhorse in

many labs," Burns says.

The study of self-assembled monolayer (SAM) films that can coat surfaces to reduce friction is a focus of work in Houston's lab, where the squarish IFM is being used to tease out the roles of chemical and mechanical configurations in affecting adhesion. Houston sounds like a biologist when he describes the properties of alkanethiol SAM films grown on gold substrates. "We can put enough stress on them to break down the gold, and they won't give up," he says.

Down the hall, ceramicist Terry Michalske points to one of the name badges worn by Sandia employees and visitors. The lab has developed a gas chromatograph with part-per-billion sensitivity that is about one-third the size of the badge. "A lot of what

we're doing is taking what we know how to do on a lab bench and miniaturizing it," he says. Accordingly, the biggest construction project in Sandia's history is scheduled to begin this year, a \$387 million fabrication plant for microelectromechanical systems.

Michalske describes current work with microfluidic devices. In particular, he and colleagues are trying to sort proteins in canals on microchips by size, charge or chemical affinities through an 'electrosomatic effect' achieved by the application of electricity. "The really interesting stuff is the molecular interactions we're programming with this method," he hints. Then he adds with a smile: "In some ways, I think the really fascinating things in nanotechnology right now are happening in beakers."

S.B.

careers and recruitment

▶ encompassing both the physical and the life sciences.

Both processes must be improved if proteomics research is to go forward, says Cameron Tropea, a lecturer in hydrodynamics and aerodynamics at the Technical University in Darmstadt, Germany. In the past, researchers thought about proteins in bulk, but now they are asking questions about physical interactions between individual proteins. This requires better temporal and spatial resolution. "There is no question that instrumentation is the limiting factor in many cases," says Tropea.

Separation anxiety

Amersham Pharmacia Biotech (APB), one of the world's leading proteomics instrumentation companies, is working on improving both separation and identification. The company's existing commercial protein-sequencing equipment works by first disrupting the cells and subsequently separating the proteins using two-dimensional (2D) gel electrophoresis. Once individual proteins have been separated into spots, lasers fire at the spots, vaporizing them. Since the electrophoresis plate holding the proteins contains a dye that absorbs light at the wavelength of the laser, the vaporized proteins can be analysed by a time-of-flight mass spectrometer.

Bjorn Nilsson, vice-president of protein area R&D at APB, believes that these 2D methods will give way to superior technologies in the future. The first stage will be to automate more of the technique. He says that semi-automatic 2D systems will speed up the rate at which proteins can be sequenced.

The company has continued to devote resources to improving this process, says Nilsson, who works at APB's centre of excellence in Uppsala, Sweden. About 300 researchers work at the centre, half of them on proteomics. Nilsson expects the work to double in the next three to four years. The advances will be aided, he says, by close links with Uppsala University, the University of Stockholm, the Royal Institute of Technology and the Karolinska Institute — one of Europe's largest medical research centres.

Mike Hayes, overall head of R&D at APB, says the company's biggest need may be for scientists working in robotics and automation technologies to help improve current processes. "The holy grail is a fully robotic system to automate 2D electrophoresis," he says.

Refining spectrometry

The mass spectrometry process can also be improved. Spectrometers can now tell what kind of a protein is present, but they are limited in telling researchers how much. At the moment, APB's equipment can resolve 10–20 peptides. Nilsson would like to see that increase by an order of magnitude.



Amersham Pharmacia Biotech's winning team: R&D head Mike Hayes (main picture), Mats Johnson (inset top) and Bjorn Nilsson.

This increase in sensitivity would also allow scientists to carry out work on single cells or even single proteins. Integration of the sample handling and sample preparation techniques will also be necessary to improve output from the mass spectrometer, according to Mats Johnson, vice-president of proteomics at APB.

Any improvements in the design of the mass spectrometer itself will require the skills of researchers in ion physics, according to Johnson. Hayes also points to the need for people who specialize in detector technologies. Improvements in proteomics instrumentation will benefit separate but related fields, such as functional and structural genomics. For example, improving robotics for proteomics could be applied to making structural genomics a more automated process, perhaps by using robotics to place samples in a synchrotron for scanning.

According to Hayes, proteomics research in general needs the skills of physical scientists with a broad background in engineering. In particular, he points to electrical and mechanical engineering as well as the detector and automation technologies mentioned above. There is a "huge opportunity for people to come into the field," he says.

Applied Biosystems (AB) also constantly recruits physicists and engineers to improve its spectrometers. But the company is

increasingly casting a wider net to refine its machines, says Dave Hicks, director of marketing.

A growing problem facing not just instrumentation but much of science — as projects continue to get bigger and more interdisciplinary — is the shortage of experienced programme managers. The ideal individual can lead a project, develop a budget and create a timeline from concept to market. These people are rare, because in addition to possessing those management skills, the ideal candidate must understand the scientific problem an instrument is designed to solve, the science behind making an instrument work, and the software that handles its data.

AB needs experts who can tweak computer databases to make them better at handling the flood of information coming off the increasingly speedy machines. These people are hard to come by, since computer and information scientists are increasingly in demand for a wider variety of disciplines.

In addition to finding jobs in instrument use or development, protein chemists with mass spectrometry experience can find jobs in technical support. It's not unusual to have PhDs taking phone calls from customers with questions. Technical service representatives also help set up the equipment and teach people in other labs how to use it.

Also, classical protein chemists with mass spectrometry experience are "getting scooped up," Hayes says — especially by companies building proteomics factories, such as AB's sister company Celera Genomics in Rockville, Maryland, or Oxford GlycoSciences in Abingdon, Oxfordshire, which has an operational proteomics factory.

Hicks expects the demand for spectrometer operators to continue as more pharmaceutical and biotech companies build such factories. Instrument companies like AB need people with such skills to help physicists and engineers develop better equipment.

"Besides supporting our customers, companies like us are competing with our customers for people who can operate these instruments," Hicks says. ■

Alok Jha is a reporter in London for *Research Fortnight*; Paul Smaglik is *Naturejobs* editor.

Web links

Nature's Genome Gateway, including a briefing on proteomics (*Nature* 402, 715–720; 1999)

♦ <http://www.nature.com/genomics/post-genomics/features.html>

Amersham Pharmacia Biotech

♦ <http://www.apbiotech.com>

Applied Biosystems

♦ <http://www.appliedbiosystems.com>

Celera Genomics

♦ <http://www.celera.com>

Oxford GlycoSciences

♦ <http://www.ogs.com>